The Use of Culture-Independent epicPCR to Determine the Host Range of Antimicrobial Resistance Genes in Manure and Manure-Fertilized Soils in Finnish Dairy Farms

Minna Maunula
Master´s thesis
University of Helsinki
April 2020

Tiedekunta – Fakultet – Faculty Faculty of Agriculture and Forestry		Koulutusohjelma – Utbildningsprogram – Degree Programme Microbiology and Microbial Biotechnology
Tekijä – Författare – Author Minna Maunula		
Työn nimi – Arbetets titel – Title The Use of Culture-Independent epicPCR to Determine the Host Range of Antimicrobial Resistance Genes in Manure and Manure-Fertilized Soils in Finnish Dairy Farms		
Oppiaine/Opintosuunta – Läroämne/Studieinriktning – Subject/Study track Microbiology		
Työn laji – Arbetets art – Level Master's Thesis	Aika – Datum – Month and year April 2020	Sivumäärä – Sidoantal – Number of pages 53

Tiivistelmä – Referat – Abstract

The use of antimicrobials in livestock production has shown to increase the abundance of antibiotic resistance genes (ARGs) in animal microbiomes. The use of manure as a fertilizer is essential in animal agriculture, however, manure application disseminates ARGs to the farm environment. In soil, the ARGs could be horizontally transferred to the environmental bacteria. Antimicrobial resistance is currently mitigated by limiting the use of antimicrobials for animals; thus, it is important to examine the ARG dynamics in countries where antimicrobial use is restricted. In addition to the antimicrobial use, also manure application rates are tightly regulated in Finnish dairy farming, offering suitable sites for examining the transmissions of ARGs in response to agricultural practices. The main aim of this study was to determine the host range of three antibiotic resistance genes by culture-independent epicPCR to understand the fate of antimicrobial resistance in agricultural environments.

The cells were extracted from manure and soil samples taken from two Finnish dairy farms. Aminoglycoside (*strB*), beta-lactam (bla_{OXA-58}) and tetracycline (*tetM*) resistance genes were linked with a phylogenetic marker gene to determine the host bacteria using epicPCR. Results were compared to the total bacterial community. In total, 664 OTU's were linked to ARGs. Antibiotic resistance genes *strB* and *tetM* shared six host genera and three genera were found to carry all the studied genes. The most common host genera for *tetM* were *Escherichia-Shigella, Sedimentibacter* and *Fibrobacter.* For bla_{OXA-58}, the most common hosts were *Sphingobacterium* and *Acinetobacter. Acinetobacter, Pseudomonas* and *Psychrobacter* carried *strB* genes in all studied samples. For the first time the host range of ARGs in manure and soil communities were determined by epicPCR, providing also valuable information for further improving comparatively new method.

Avainsanat – Nyckelord – Keywords
Antimicrobial resistance genes, Agroecosystem, Culture-Independent, Environmental microbiology

Ohjaaja tai ohjaajat – Handledare – Supervisor or supervisors
Johanna Muurinen, Katariina Pärnänen, Jenni Hultman & Marko Virta

Säilytyspaikka – Förvaringsställe – Where deposited
HELDA - Helsingin yliopiston digitaalinen arkisto / HELDA - Helsingfors universitets digitala publikationsarkiv /HELDA - Digital Repository of the University of Helsinki

Muita tietoja – Övriga uppgifter – Additional information
Funded by: Academy of Finland, Maj and Tor Nessling Foundation

Tiedekunta – Fakultet – Faculty Maatalous-metsätieteellinen tiedekunta	Koulutusohjelma – Utbildningsprogram – Degree Programme Mikrobiologian ja Mikrobibiotekniikan maisteriohjelma	
Tekijä – Författare – Author Minna Maunula		
Työn nimi – Arbetets titel – Title Antibioottiresistenssigeenien isäntäbakteerikirjon määrittäminen viljelyvapaalla epicPCR menetelmällä suomalaisilla maitotiloilla lannassa ja lannoitetussa peltomaassa		
Oppiaine/Opintosuunta – Läroämne/Studieinriktning – Subject/Study track Mikrobiologia		
Työn laji – Arbetets art – Level Maisterintutkielma	Aika – Datum – Month and year Huhtikuu 2020	Sivumäärä – Sidoantal – Number of pages 53

Tiivistelmä – Referat – Abstract

Eläintuotannossa käytettävien mikrobilääkkeiden on osoitettu lisäävän antibioottiresistenssigeenien esiintyvyyttä tuotantoeläinten mikrobiomeissa. Lannan käyttö lannoitteena on olennaista maataloudessa, mutta se levittää antibioottiresistenssigeenejä maatilaympäristöön. Maaperässä resistenssigeenit voivat siirtyä horisontaalisesti ympäristöbakteereille. Mikrobilääkeresistenssin leviämistä pyritään hillitsemään rajoittamalla mikrobilääkkeiden käyttöä eläimillä. Sen vuoksi ARG dynamiikkaa on tärkeää tutkia maissa, joissa mikrobilääkkeiden käyttö on rajoitettua. Mikrobilääkkeiden käytön lisäksi myös lannan levitysmäärät ovat tiukasti säänneltyjä suomalaisessa maitokarjataloudessa, minkä vuoksi suomalaiset maitotilat ovat sopivia antibioottiresistenssin leviämisen tutkimiseen maatalouskäytäntöjen funktiona. Tämän tutkimuksen päätavoitteena oli määrittää mitkä bakteerit kantavat kolmea antibioottiresistenssigeeniä ilman kasvatusta käyttäen epicPCR-menetelmää ja siten edistää ymmärrystä mikrobilääkeresistenssistä maatalousympäristöissä.

Lanta- ja maanäytteitä kerättiin kahdelta suomalaiselta maitotilalta bakteerisolujen eristystä varten. Aminoglykosidi (*strB*), beeta-laktaami (bla_{OXA-58}) ja tetrasykliini (*tetM*) -resistenssigeenit yhdistettiin fylogeneettisella tunnistusgeeniä käyttäen isäntäbakteereihin epicPCR menetelmää käyttämällä. Tuloksia verrattiin koko bakteeriyhteisöön. Kaikkiaan 664 OTU:a yhdistettiin tutkittuihin geeneihin. *strB* ja *tetM* jakoivat kuusi isäntäsukua ja kolmen suvun todettiin kantavan kaikkia tutkittuja geenejä. *tetM*:n yleisimmät isäntäsuvut olivat *Escherichia-Shigella, Sedimentibacter* ja *Fibrobacter.* bla_{OXA-58} geenin yleisimmät isännät olivat *Sphingobacterium* ja *Acinetobacter. Acinetobacter, Pseudomonas* ja *Psychrobacter* kantoivat *strB* geeniä kaikissa tutkituissa näytteissä. EpicPCR:ää käytettiin ensimmäistä kertaa määrittämään lanta- ja maaperäyhteisöjen antibioottiresistenssigeenien isäntäkirjoa ja tämä työ tarjosi samalla arvokasta tietoa suhteellisen uuden menetelmän edelleen kehittämiseksi.

Avainsanat – Nyckelord – Keywords
Antibioottiresistenssigeenit, Maatalouden ekosysteemi, Viljelyvapaa menetelmä, Ympäristömikrobiologia

Ohjaaja tai ohjaajat – Handledare – Supervisor or supervisors
Johanna Muurinen, Katariina Pärnänen, Jenni Hultman & Marko Virta

Säilytyspaikka – Förvaringsställe – Where deposited
HELDA - Helsingin yliopiston digitaalinen arkisto / HELDA - Helsingfors universitets digitala publikationsarkiv /HELDA - Digital Repository of the University of Helsinki

Muita tietoja – Övriga uppgifter – Additional information
Rahoitus: Academy of Finland, Maj ja Tor Nesslingin Säätiö

Introduction

Antibiotic resistance is one of the ten most important threats to global health in 2019 (WHO, 2019). Animal agriculture is one of the biggest sectors using antibiotics. It has been estimated that in 2010 the consumption of antibiotics in livestock was 63 151 tons globally and that the consumption will increase by 67% by 2030 (van Boeckel et al., 2015). Up to 75–90% of the used antimicrobials in livestock are excreted in feces and urine as unmetabolized molecules (FAO, 2019). The use of antibiotics both in humans and in agriculture has driven the speed of evolution of antibiotic resistance bacteria (ARBs) faster than ever (Gillings & Stokes, 2012). ARBs can be transferred to humans from the farm environment (Graham et al., 2009), by direct contact with animals (Smith et al., 2013) or by consumption of food products (Witte, 2000; Price et al., 2005). It has been shown that gulls and geese nesting close to production animal farms have more resistant *Escherichia coli* than birds that are connected to more pristine water bodies (Cole et al., 2015; Dolejska et al., 2007). Thus, agricultural environment is a potential route for the transfer the resistance into wildlife (Allen et al., 2010). The movement of ARB between humans, animals and the environment highlight the importance of One Health approach to tackle antimicrobial resistance, which states that human and animal health and environment are interconnected (One Health Global Network, 2012).

While it has been demonstrated that animal husbandry increases the prevalence of antibiotic resistance genes (ARGs) in the farm environment (Knapp et al., 2010), the environment is also rich in bacteria that are intrinsically resistant to many clinically used antibiotics. Some of the environmental bacteria are naturally capable of producing antibiotics (Martinez, 2008; Wright, 2010). In fact, many bacteria may have acquired resistance genes against some clinically important antibiotics already a long time before modern antibiotic era (Petrova et al., 2009) and some of the clinically emerging resistance genes may have environmental origin (Wright, 2010; Surette & Wright, 2017). However, it is unclear under which conditions the resistance genes in the environment are mobilizable.

The use of antibiotics in agriculture causes selective pressure for resistance in gastrointestinal bacteria and bacteria in stored manure (Chee-Sanford et al., 2009). Bacteria need to adapt by developing resistance against the antibiotics through mutations or by acquiring resistance genes horizontally (Martinez, 2008; Allen et al., 2010; Wellington et al.,

2013). Horizontal gene transfer (HGT) and mobilization of ARGs are associated with stress response, which can be caused by antibiotic use or other pollutants in the environment (Gillings & Stokes, 2012; Wellington et al., 2013). HGT can happen with the help of plasmids or mobile genetic elements (MGEs) including transposons and integrons (Furuya et al., 2006; Allen et al., 2010; Wellington et al., 2013) even between distant relatives (Tamminen et al., 2012). Animal manure disseminates ARGs to the farm environments (Ruuskanen et al., 2016) and is considered to be a potential route to transfer resistance genes into human and animal pathogens (Baquero et al., 2008).

The use of antibiotics for production animals in Finland is tightly restricted and was one of lowest amongst the 30 studied European countries in 2016 (European Medicines Agency, 2016). In Finland, dairy cattle are predominantly grass-fed, and the grass is fertilized with manure that is stored over the winter in silos or lagoons. Manure application is regulated by law, and in practice, manure can only be applied from the beginning of April until the end of October/early November (Finlex, 2000). Muurinen et al. (2017) explored the influence of manure storage and land application on the relative abundance of ARGs and MGEs in Finnish dairy and swine farms, and more than two hundred ARGs and MGEs were found. The ARGs were most abundant in manure and storage increased the abundance of most of the ARGs. Many of the ARGs that were abundant right after manure-fertilization were not detected in soil later in the growing season, but few of the ARGs became more abundant in soil over time. The results indicated that despite the manure application and antibiotic use are restricted in Finland, the agroecosystems are rich in ARGs and MGEs and the long-term storage of the manure increases the load of ARGs and MGEs that are disseminated to the environment among manure fertilization. Also, due to agricultural practices, the ARBs and ARGs circulate in the farm environment. The ARGs can disseminate from fertilized soil via harvested feed to cattle microbiome, from the gastrointestinal tract to manure and from manure back to soil (Figure 1). To investigate the potential risks caused by ARG transmission in this circle, host range identification is needed. More precisely, obtaining information of the carrier's potential pathogenicity and their capability to transfer these genes horizontally will increase our understanding on antibiotic resistance in human impacted environments.

When a host bacterium of certain gene needs to be examined, so far epicPCR is one of the few methods that can link a gene of interest to its unculturable host bacterium. Compared to multitude existing methods using metagenomic approaches, epicPCR narrows down the

amount of information, only offering sequence data on the carriers of the selected genes, which simplifies the analysis. EpicPCR method is based on a single-cell technique where one living bacterial cell is captured inside a polyacrylamide bead. The beads have a half diffusive membrane that has a pore size large enough to let PCR reagents inside the bead but small enough for keeping the cell inside the bead (Spencer et al., 2016). In the first step of epicPCR, the phylogenetic marker gene, 16S rRNA gene, is linked with a linker primer to a target gene in a fusion PCR. Next, blocking PCR is run to prevent amplification of unfused products, followed by nested PCR to get higher concentration of the fused product.

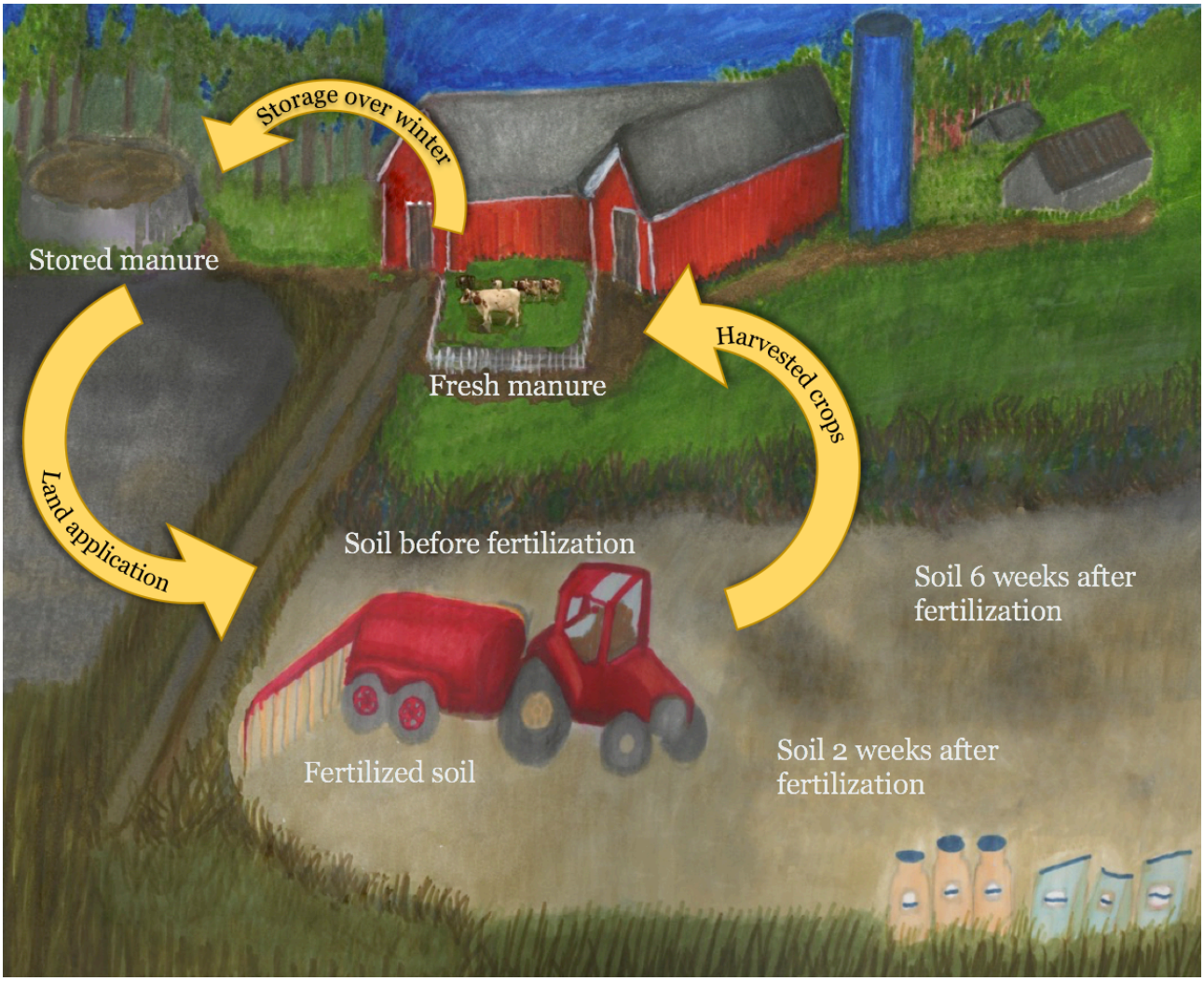

Figure 1. Modified from Muurinen et al. (2017). Yellow arrows show potential dissemination route of ARBs and ARGs in Finnish agricultural practises. Sample types used in this and the study of Muurinen et al. (2017) are in marked in white.

Three antibiotic resistance genes were chosen to this study based on the study by Muurinen et al. (2017). Resistance gene *tetM* confers resistance to tetracyclines and has been found in both gram-positive and gram-negative species (Chopra et al., 2001). *tetM* is commonly located in a conjugative mobile genetic element, such as Tn916-like transposons, which encode their own integration and transfer functions (Chopra et al., 2001). bla_{OXA-58} is a carbapenem-hydrolyzing oxacillinase gene. Carbapenems are so called last resort antibiotics that are used for treating infections caused by multidrug-resistant gram-negative pathogens (Leski et al., 2013). bla_{OXA-58} has found in plasmids and commonly locates close to insertion sequence ISAba3 (Poirel & Nordmann, 2006). Aminoglycoside resistance gene *strB* is also associated with mobile genetic elements. It can occur as a *strA*-*strB* gene pair and is linked with integrons and transposons in small nonconjugative broad-host-range plasmids as well as in large conjugative plasmids (Chiou & Jones, 1993; Sundin & Bender, 1996a; Sundin, 2002). Since all of the aforementioned genes are linked to MGEs, they have potential to spread through HGT, which could be induced with antimicrobial use and manure storage (Muurinen et al., 2017).

Aims of this study was to determine the host range of three ARGs from agricultural soils and manure in two Finnish dairy farms (Figure 1) without culturing and compare the hosts to the total microbial community and their abundances in the samples. The host range of *strB*, *tetM* and bla_{OXA-58} were determined using culture independent epicPCR for cells extracted from manure and manure-fertilized soils. The total microbial community was analyzed with 16S rRNA gene sequencing. epicPCR has been previously used for environmental samples (Spencer et al., 2016; Hultman et al., 2018; Qin et al., 2019), but not in soil and manure. This study aimed also to examine the use and potential of epicPCR in these kinds of samples.

Methods

Sample collection and pre-treatment

Samples were collected from two Finnish dairy farms (Farm 1 and Farm 2) located in Southern Finland approximately 100 km from Helsinki between April and June in 2017. Farm 1 has approximately 400 animals and Farm 2 has 240 animals. Manure was stored in Farm 1 in a concrete silo and in a lagoon in Farm 2. The following samples were collected: Fresh and stored manure, soil before manure fertilization, soil after manure fertilization, soil two weeks after manure fertilization and soil six weeks after manure fertilization. Sample handling is described in Ruuskanen et al. (2016). For epicPCR, 5 g of soil or manure was measured into a 15 ml falcon tube and the samples were suspended by manual shaking with ~30% v/v glycerol after which the samples were flash frozen with dry ice containing 99.7% v/v ethanol and stored at -80°C.

Three biological replicates were processed from each sample. Samples were pre-treated with sonication in order to separate the cells from soil and manure particles. Glycerol was removed by centrifugation at 11 700 RCF at 4°C for 2 min and the bacterial cells were suspended into 700 μl of 1x phosphate buffer solution (PBS). A sterile wooden stick was used to help in suspending the pellet before transfer into Covaris tubes (13 x 65 mm with a Covaris cap 13 x 65 mm). Tubes were compatible with S220 Focused-ultrasonicators (Covaris, USA) used with Recirculator Model 13270-120 (VWR™) to keep the samples at 4°C temperature. Samples were vortexed gently before placing the tube into the sonicator. Settings for sonication were as follows for soil; duty cycle 1%, Intensity 0,1 and cycles per burst 100 and for manure; duty cycle 2%, intensity 1 and cycles per burst 200. Samples were kept on ice between steps.

Two milliliters of 1.3 g ml^{-1} freshly prepared nonionic density gradient medium (Histodenz™, Sigma-Aldrich, China) solution was used to separate cells from the sample matrix by non-ionic gradation. 350 µl of sonicated and vortexed soil sample was transferred on top of the Histodenz solution and centrifuged for 20 min at 5000 RCF at 4°C. Top and middle phases were transferred into a new microcentrifuge tube. Manure samples were strained trough a 35 µm cell strainer (Falcon, USA) before transfer. Same steps were repeated to the remaining sample volume. After cell collection, samples were centrifuged

for 10 min at 13 000 RCF at 4°C. Supernatant was removed and the pellet was suspended in 50 µl of 1x PBS solution and stored overnight at 4°C.

epicPCR

The epicPCR protocol was carried out following Spencer et al. (2016) with some modifications. The cell suspensions from soil samples were diluted with PCR-grade H_2O before polymerization with the dilution factors ranging from 1:5, 1:10, 1:20, 1:30, 1:50 to 1:100 based on approximated estimations of the density of cells in the sample suspension that were obtained by fluorescence microscoping (Zeiss Axioskop 2 plus, Oberkochen, Germany) the pretreated samples. For this, a subsample of the diluted cells was stained with SybrGreenII (Thermo Fisher Scientific, Waltham, USA) and the cells were visualized using UV light. Before the polymerization step of epicPCR, 30 µl of the diluted cells were combined with 100 µl PCR-grade H_2O, 100 µl of 30% w/v Acrylamide/BIS solution, 29:1 (Bio Rad, China) and 25 µl of 10% w/v APS (Sigma, USA). The suspension was vortexed in a 2 ml safety-lock microtube. 600 µl of STT emulsion oil (4.5% Span 80 (Sigma, USA), 0.4% Tween 80 (Sigma-Aldrich, USA), 0.05% Triton X-100 (AppliChem, Germany), v/v in Mineral oil (Sigma, USA)) was added and the samples were vortexed vertically in 3000 RCF for 30 s. 25 µl of 100% v/v TEMED (Sigma-Aldrich, China) was added to the sample and suspension was vortexed as in the previous step. Tubes were left to polymerize for 90 minutes in room temperature.

After polymerization, the oil was removed by adding 800 µl of diethyl ether (50:50 diethyl ether/MQ-H_2O) (VWR Chemicals, EC), into the tube. After addition of ether the tube was immediately inverted and flicked so that a visible cloudy-like precipitate was formed. Diethyl ether was removed, and the resulting precipitate was washed with 1 ml of sterile MQ-H_2O by flicking and inverting technique. The suspension was centrifuged 30 s 12 000 RCF. Washing steps were repeated approximately 10 times or until all of the oil was removed, and the liquid phase was transparent. Finally, remaining water was removed, and the beads were resuspended in 1 ml of 1 x TK buffer (20 mM Tris-HCl (pH 7.5), 60 mM KCl) after straining trough a 35 µm cell strainer (Falcon, USA) and the cells were transferred into a new tube. Samples were stored overnight at 4°C to let the beads settle. On the following day the beads were checked by microscoping (Zeiss Axioskop 2 plus, Oberkochen,

Germany) to confirm that one bead contained only one cell (Supplementary Figure S1). Cells were stained with SybrGreenII (Thermo Fisher Scientific, Waltham, USA).

Before fusion PCR, 46.5 µl of sample, 53.5 µl of PCR-mastermix and 900 µl ABIL emulsion oil (4% ABIL EM-90 (Evonik, Germany), 0.05% Triton X-100 (AppliChem, Germany), v/v in Mineral oil (Sigma, USA)) was mixed in a 2 ml safe-lock Eppendorf tube containing 4 x 2 mm glass beads by vortexing vertically with maximum speed for 1 min. PCR amplification for the target genes and for the 16S rRNA gene target region was performed with 10 µM R2 and F1 primers and 1 µM R1-F2´primer (Table 1). The PCR reaction was run in volume of 50 µl in 1 x GC buffer with 1 mM $MgCl_2$, 0.2 U μl^{-1} Phusion Hot Start Flex polymerase (New England Biolabs, USA) and 0.25 mM dNTPs (Bionordika, Finland). The PCR conditions were as follows: beginning temperature of 80°C for 10 s, initial denaturation at 94°C for 30 s, followed by 32 cycles of denaturation at 94°C for 5 s, annealing at 55°C for 30 s and extension at 72°C for 30 s, after which final extension was performed at 72°C for 5 min before cooling to 4°C.

Immediately after fusion PCR, ABIL emulsion phase was separated and purified with 1 ml of water-saturated diethyl ether (VWR Chemicals, EC) and vortexed and centrifuged for 1 min in 13 000 RCF. After centrifugation, the upper phase was discarded and 50 µl of PCR-H_2O and 1 ml water saturated ethyl acetate (Sigma-Aldrich, USA) was added. This was followed by centrifugation done as in the previous step. Two more diethyl ether extractions were performed and finally after upper phase removal, the Eppendorf tubes were left open for approximately 5 min, in order to let the remaining diethyl ether to evaporate. The phase containing the DNA was transferred into a new tube and the PCR product was purified with Monarch PCR & DNA Cleanup kit (New England Biolabs, USA) according to manufacturer's protocol.

The blocking PCR was run in a volume of 25 µl in 1 x GC buffer (Thermo Fisher Scientific), with the concentrations of 3.2 µM Block F and 3.2 µM Block R primers (Table 1), 0.2 mM dNTPs (Bionordika, Finland) and 0.02 U μl^{-1} of Phusion High-Fidelity polymerase (Thermo Fisher Scientific). The PCR conditions were as follows: beginning temperature at 98°C for 30 s, followed by 30 cycles of denaturation at 98°C for 10 s, annealing at 55°C for 30 s and extension at 72°C 30 s, after which a final extension was performed at 72°C for 5 min before cooling to 4°C. The PCR products were purified with Monarch PCR & DNA Cleanup kit (New England Biolabs, USA) according to manufacturer's protocol.

Nested PCR was performed for all of the samples. Negative controls were included for all of the studied genes. The nested PCR was performed with the same master mix than blocking PCR, but with different primer concentrations. The blocking primer concentration was 0.32 µM and primers F3 and R3 were 0.3 µM (Table 1). PCR conditions were as in the blocking PCR, but 35 cycles instead of 30 were run. In total, four 25 µl reactions were performed for each sample. All PCR reactions were performed with a C1000 Touch Thermal Cycler (Bio-Rad Laboratories, Hercules, USA) and the gel electrophoresis for the epicPCR products performed with the E-Gel iBase Power System (Thermo Fisher Scientific) using E-Gel EX Gel, 2% (Invitrogen) agarose gels. A 1 Kb Plus DNA Ladder (Thermo Fisher Scientific) was used to estimate the size of the final epicPCR amplification products. All the samples were sequenced twice due to sequencing errors, which are discussed in results and discussion section. Sequencing was done with the Illumina Miseq platform at the Institute of Biotechnology, University of Helsinki, Finland and the resulting sequencing libraries were combined for analysis.

DNA extraction and 16S rRNA gene PCR

DNA was extracted and diluted into 100 µl with DNeasy PowerLyzer PowerSoil Kit (Qiagen, Germany) according to manufacturer's protocol. DNA concentration was measured with Qubit Broad-Range Assay Kit (Thermo Fisher Scientific, Waltham, USA). The V3-V4 region of the 16S rRNA gene was amplified in a reaction volume of 25 µl in 1 x GC buffer, 2.5% v/v DMSO (Thermo Fisher Scientific), with 0.2 µM 341F1-4 and 0.2 µM 785R1-4 primers (Table 1), which both contained the Illumina TruSeq adapters in 5' ends, 0.2 mM dNTPs (Bionordika, Finland) and 0.02 U μl^{-1} of Phusion High-Fidelity polymerase (Thermo Fisher Scientific). Only one replicate from each sample was used for 16S rRNA gene amplification. The PCR conditions were as follows: Initial denaturation at 98°C for 30 s, followed by 14 cycles with denaturation at 98°C for 10 s, annealing at 60°C for 30 s and extension at 72°C for 10 s, after which a final extension was performed at 72°C for 5 min before cooling to 4°C. PCR products were purified with Monarch PCR & DNA Cleanup kit (New England Biolabs, USA) according to manufacturer's protocol. All samples were sequenced on the Illumina Miseq platform at the Institute of Biotechnology, University of Helsinki, Finland.

Table 1. Information of the primers used in this study. Modified from Hultman et al. (2018)

Primer name	sequence 5'-3'	Target gene and primer name	Reference
F1_tetM	CATCATAGACACGCCAGGACA	*tetM* epic forward	Karkman et al. (2016)
R1-F2'_tetM	**GWATTACCGCGGCKGCTG**CTGTTTG ATTACAATTTCCGC	*tetM* epic linker	Tamminen et al. (2011)
F3_tetM_TS	**ATCTACACTCTTTCCCTACACGACG CTCTTCCGATCT**GCAATTCT ACTGATTTCTGC	*tetM* epic nested	Tamminen et al. (2011)
F1_blaOXA-58	ACAGGCACTGTAGATGCTTG	bla_{OXA-58} epic forward	Karkman et al. (2016)
R1-F2'_blaOXA-58	**GWATTACCGCGGCKGCTG**TGTGCT GAGCATAGTATGAG	bla_{OXA-58} epic linker	Karkman et al. (2016)
F3_blaOXA-58 _TS	**ATCTACACTCTTTCCCTACACGACG CTCTTCCGATCT**TCGGTCTA AATGCGTGCCAT	bla_{OXA-58} epic nested	Karkman et al. (2016)
F1_strB	CTAATGGCGAAGCTGTATG	*strB* epic forward	This thesis
R1-F2'_strB	**GWA TTA CCG CGG CKG CTG** GTG GAC GTA GTC AGT TTG AC	*strB* epic linker	This thesis
F3_strB_TS	**ATCTACACTCTTTCCCTACACGACG CTCTTCCGATCT**GTATGCCGCATCTG AGGAAC	*strB* epic nested	This thesis
Illum_785R_1 (R3)	**GTGACTGGAGTTCAGACGTGTGCTC TTCCGATCT**GACTACHVGGGTATCTA ATCC	16S rRNA gene reverse	Herlemann et al. (2011)
Illum_785R_2 (R3)	**GTGACTGGAGTTCAGACGTGTGCTC TTCCGATCT**aGACTACHVGGGTATCT AATCC	16S rRNA gene reverse	Herlemann et al. (2011)
Illum_785R_3 (R3)	**GTGACTGGAGTTCAGACGTGTGCTC TTCCGATCT**tctGACTACHVGGGTATC TAATCC	16S rRNA gene reverse	Herlemann et al. (2011)
Illum_785R_4 (R3)	**GTGACTGGAGTTCAGACGTGTGCTC TTCCGATCT**ctgagtgGACTACHVGGGT ATCTAATCC	16S rRNA gene reverse	Herlemann et al. (2011)
R2_1492R	GGTTACCTTGTTACGACTT	16S rRNA gene reverse	Lane, (1991)
U519F_block (Block F)	TTTTTTTCAGCMGCCGCGGTAATWC/ 3SpC3/	partial fusion products	Spencer et al. (2016)
U519R_block (Block R)	TTTTTTTGWATTACCGCGGCKGCTG/ 3SpC3/	partial fusion products	Spencer et al. (2016)
Illum_341F_1	**ATCTACACTCTTTCCCTACACGACG CTCTTCCGATCT**CCTACGGGNGGCW GCAG	16S rRNA gene forward	Herlemann et al. (2011)

Illum_341F_2	**ATCTACACTCTTTCCCTACACGACG CTCTTCCGATCT**gtCCTACGGGNGGC WGCAG	16S rRNA gene forward	Herlemann et al. (2011)
Illum_341F_3	**ATCTACACTCTTTCCCTACACGACG CTCTTCCGATCT**agagCCTACGGGNG GCWGCAG	16S rRNA gene forward	Herlemann et al. (2011)
Illum_341F_4	**ATCTACACTCTTTCCCTACACGACG CTCTTCCGATCT**tagtgtCCTACGGGNG GCWGCAG	16S rRNA gene forward	Herlemann et al. (2011)

[1]Lower case nucleotides in primers Illum_341 and Illum_785 are differing from each other's by nucleotides and resulting more variability in the end product by mixing in sequencing.

[2]Nucleotides in bold highlight the 16S rRNA gene sequence (in R1_F2') or the short Illumina TruSeq adapter (in nested primers).

16S rRNA gene sequence analysis

Bacterial community analysis was done following a pipeline created by Jenni Hultman (https://github.com/jjholsa/MMB-117). Adapters were removed using cutadapt v.1.10 (Martin, 2011) with the cutadapt command and -m 1, -e 0.2, -O 15, -g, -G, -q 25 and -p parameters. FastQC v.0.11.8 (Andrews, 2010) was used to analyse the quality of the trimmed 16s rRNA amplicons with command fastqc -t 4. Reads were joined using Pear v.0.9.6 (Zhang et al., 2014) with command pear and default options. Length distribution of the reads was checked with Prinseq v.0.20.4-4 (Schmieder & Edwars, 2011) using command prinseq-lite.pl. Reads were quality trimmed using USEARCH v.10 (Edgar, 2013) with fastq_filter command and -threads 2, -fastq maxee 1, -fastq minlen 350 parameters. Unique sequences were identified with the UPARSE pipeline (Edgar, 2013) with fastx uniques command. Chimeras were removed, and reads were mapped to reference sequence to make operational taxonomic units (OTUs) (with the default 97%) with the -cluster_otus command. OTU-table and the abundance of OTUs were acquired with -otutab command. Finally, the taxonomic classification of OTUs was done with RDP naïve Bayesian Classifier Mothur v.1.42.0 (Schloss et al., 2009) against the Silva 132 database (Quast et al., 2013) with command classify.seqs with cutoff=60 and processors=4 parameters.

epicPCR data analysis

The analysis of epicPCR amplicon sequences was done following epicPCR analysis guide created by Katariina Pärnänen (https://github.com/KatariinaParnanen/epicPCR_analysis). The quality of the paired end reads was analyzed with FastQC v.0.11.8 (Andrew, 2010). Adapters were removed with cutadapt v.1.10 (Martin, 2011) using cutadapt command and with -p parameters. Reads were joined with Pear v.0.9.6 (Zhang et al., 2014) with pear command with -y 150M and -j 2 parameters. Both sequence runs were combined and analysed at the same time. The quality of the merged reads was analyzed again with FastQC. The removal of 16S end primer and quality filtering was done with cutadapt command with –max n=5, -q 20, -m 350, -M 600 parameters. The 16S rRNA gene sequences and ARG sequences were split into different files based on primer sequences using cutadapt. First cutadapt command was ran to all filtered samples extraction reads starting with the target gene forward primer F3_ with -g -O 10, -e 0.2 parameters. The 16S region was obtained using the forward 16S primer U519F_block and target gene region extracted using the reverse part from linker primer R1-F2'_ with cutadapt command with -O 15 parameter. Singletons were removed from the 16S rRNA sequences using vsearch v.2.6.0 (Rognes et al., 2016) with derep_fulllength command and –minuniquesize 2 parameters. Sequences were clustered with cluster_fast command with –id 0.97, --centroids, --relabel OTU, --uc parameters. Reads were mapped back with usearch_global command with –db, --strand plus, --id 0.97, --uc options. The taxonomic classification of OTUs was done with RDP naïve Bayesian Classifier Mothur v.1.40.4 (Schloss et al., 2009) against the Silva 132 database (Quast et al., 2013) with command classify.seqs with cutoff=60, processors=1 and probs=F options. The ARGs were annotated with BlastN (Altschul et al., 1990) to verify the correct amplification of the target genes. The size of the fused epicPCR products varied between 480-545 bp. The quality of first sequencing round was low and the sequence quality of the reads dropped after 100 bp. Samples aligned poorly with Pear and after filtering the reads with Cutadapt by the correct size, less than 1% of the reads passed. Samples were re-sequenced to get more reads and treated with more aggressive size selection at the sequencing facility.

Statistical analysis

Statistical analysis and visualization of the 16s rRNA gene sequencing and epicPCR results was done using R version 3.6.0 (R Core team, 2017) and Rstudio (Rstudio team, 2015). Phyloseq v.1.28.0 (McMurdie & Holmes, 2013) was used to combine the OTU and taxonomy tables and for calculating the beta diversity in the bacterial communities using the Bray-Curtis dissimilarity index. The community compositions between farms and samples were compared with permutational multivariate analysis of variance in the vegan package v.2.5-6 (Oksanen et al., 2017) using *adonis* function. Beta-diversity was determined using both Bray-Curtis dissimilarity index and Jaccard distance index and 9999 permutations. The 16S rRNA gene counts were normalized to the library sizes. The results were visualized using ggplot2 v.3.2.1 (Wickham, 2016) The analysis of the 16s rRNA gene data in R was done following a pipeline created by Igor Pessi (https://github.com/igorspp/MMB-114/blob/master/MMB-117.md). Contaminants in epicPCR were removed by Decontam v.127.0.0.1 (Davis et al., 2017) by prevalence. Threshold 0.5 for carriers of *tetM* and threshold 0.1 for carriers of bla_{OXA-58} and *strB* were used, since more reads was found from negative controls of *tetM*. If the ARG was present in the same OTU in 2/3 biological replicate, the result was considered to be reliable. The analysis and visualization was done based on epicPCR analysis in Github created by Katariina Pärnänen (https://github.com/KatariinaParnanen/epicPCR_analysis/blob/master/epic_R.html). More detailed version of used commands is available in the Supplementary material.

Results

Bacterial community

The total bacterial communities in fresh and stored manure, and in unfertilized and fertilized soils between farms were observed to be similar. After quality filtering, a total of 63 337 reads and 622 OTUs were obtained from 16S rRNA gene sequence data. The profiles of most abundant genera were dissimilar between soil and manure, but similar among the sample types. Three bacterial orders were dominant in the communities found from the manure. Clostridiales (42.1%), Bacteroidales (23.9%) and Pseudomonadales (21.3%) were the most dominant orders in the manure followed by Lactobacillales (6.6%) and Corynebacteriales (1.2%). The bacterial community in the soil was more diverse than in the manure. The most abundant order in soil was Rhizobiales (11.4%), followed by Micrococcales (6.5%), Gemmatimonadales (5.8%), Gaiellales (5.8%) and Sphingomonadales (5.3%). Bacterial communities between soil and manure were observed to be significantly different and explained by the sample type (PERMANOVA with Bray-Curtis dissimilarity index; $P = 0.0001$, $R^2 = 0.62$ and Jaccard distance index; $P = 0.0004$, $R^2 = 0.69$). Visualization of Beta-diversities of the microbial communities showed that the sample type had an effect on the microbial communities between manure and soil (PERMANOVA with Bray-Curtis dissimilarity, $P = 0.03$, and $R^2 = 0.15$) (Figure 2). Soil samples were clustered more closely together compared to manure samples. No significant difference was observed between the farms, even though the community in fresh manure samples in different farms was found to be dissimilar with each other (Figure 2).

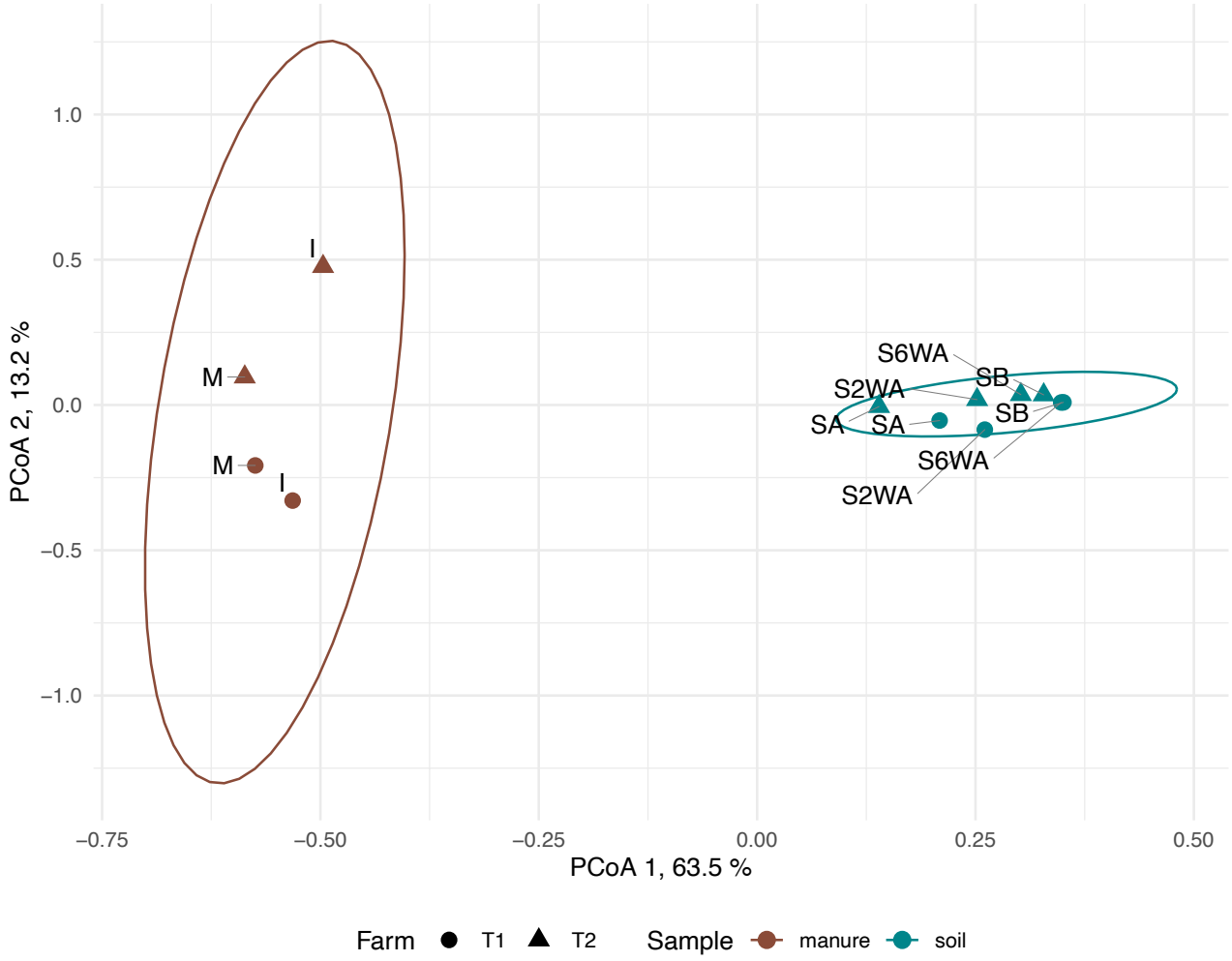

Figure 2. Principal coordinate analysis (PCoA) plot from the relative abundance of OTUs between the sample types in Farm 1 and Farm 2 with the Bray-Curtis dissimilarity. 16S rRNA gene sequences were normalized to proportions. The circles around the sample types are drawn with confidence level of 0.95. Sample types are significantly different if ellipses are not overlapping. Soil samples from Farm 1 and Farm 2 are soil before (SB), soil after (SA), soil 2 weeks after (S2WA) and soil 6 weeks after (S6WA) fertilization and marked in turquoise. Manure samples from Farm 1 and Farm 2 are fresh (I) and stored manure (M) and marked in brown. Circle shaped symbol stands for Farm 1 and triangle for Farm 2.

At the order level, manure samples had similar bacterial community composition on both farms, but the abundances varied in fresh and stored manure (Figure 3). Pseudomonadales were the most abundant order in fresh manure Farm 1, when Clostridiales were the most abundant order in fresh manure Farm 2. During manure storage, the abundance of the three most abundant orders evened out. In soil, the abundance and composition of bacterial communities was similar in both farms, both in order (Figure 3) and genus level (Figure 4). At the genus level in the fresh manure community, the genera were having very different

pattern between farms (Figure 5). Stored manure communities had very similar composition regardless of the farm (Figure 5).

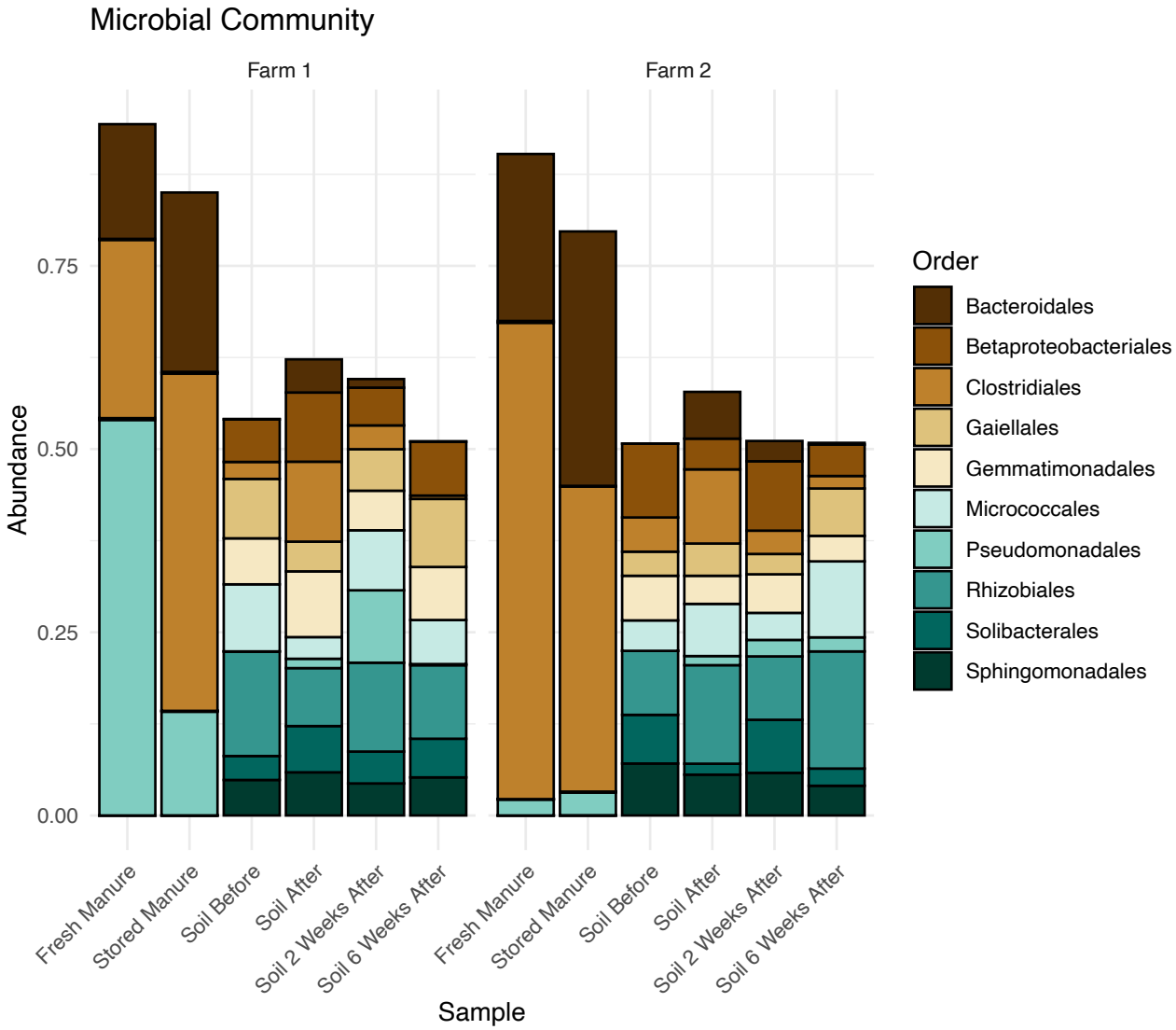

Figure 3. Relative abundance bar plot of the 16S rRNA gene sequences at the order level in different sample types in Farm 1 (left) and Farm 2 (right). The 16S rRNA gene counts were normalized to the library sizes. Soil samples are taken in different time points before and after manure fertilization.

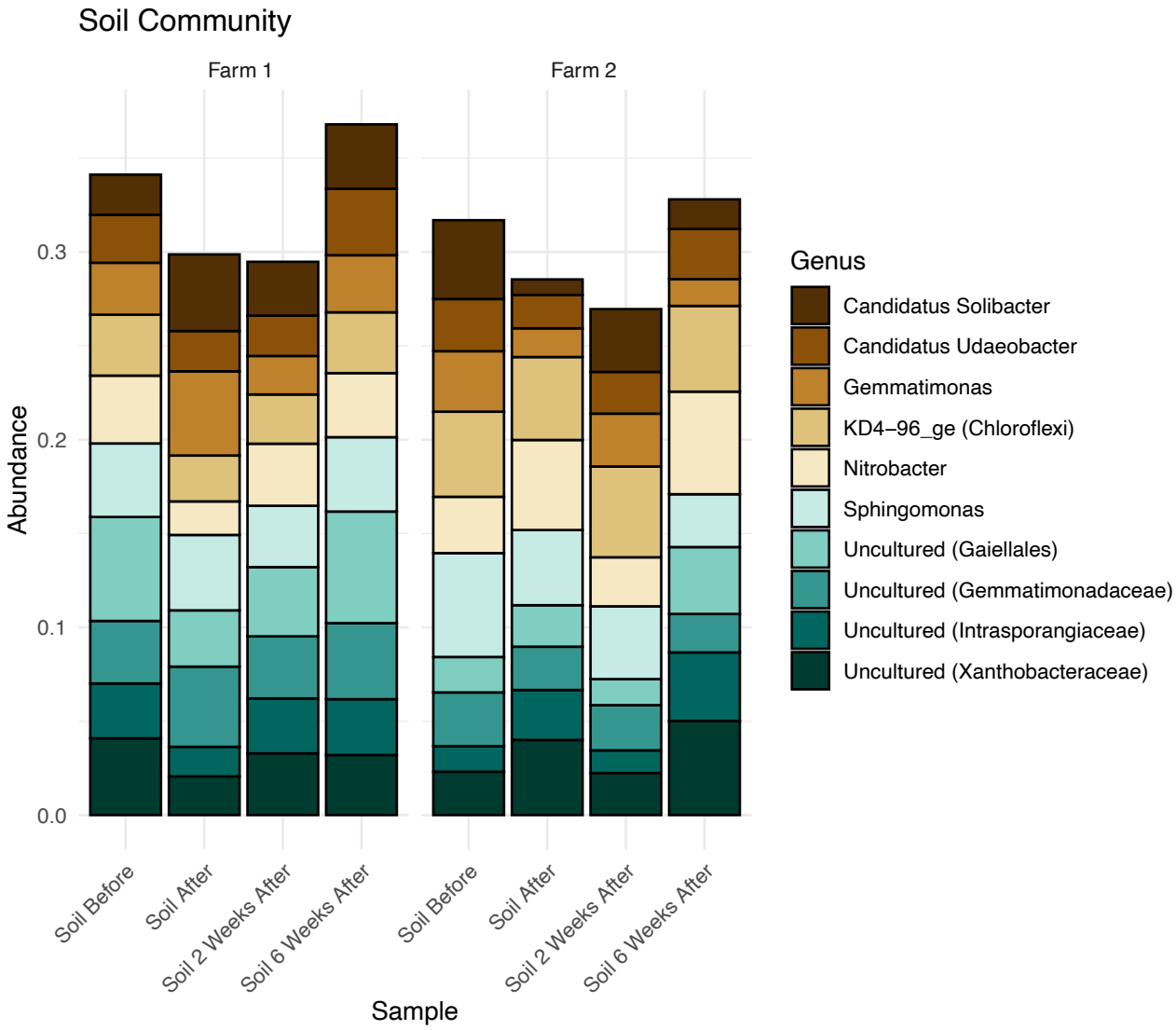

Figure 4. Relative abundance bar plot of the 16S rRNA gene sequences at the genus level in different soil sample types in Farm 1 (left) and Farm 2 (right).

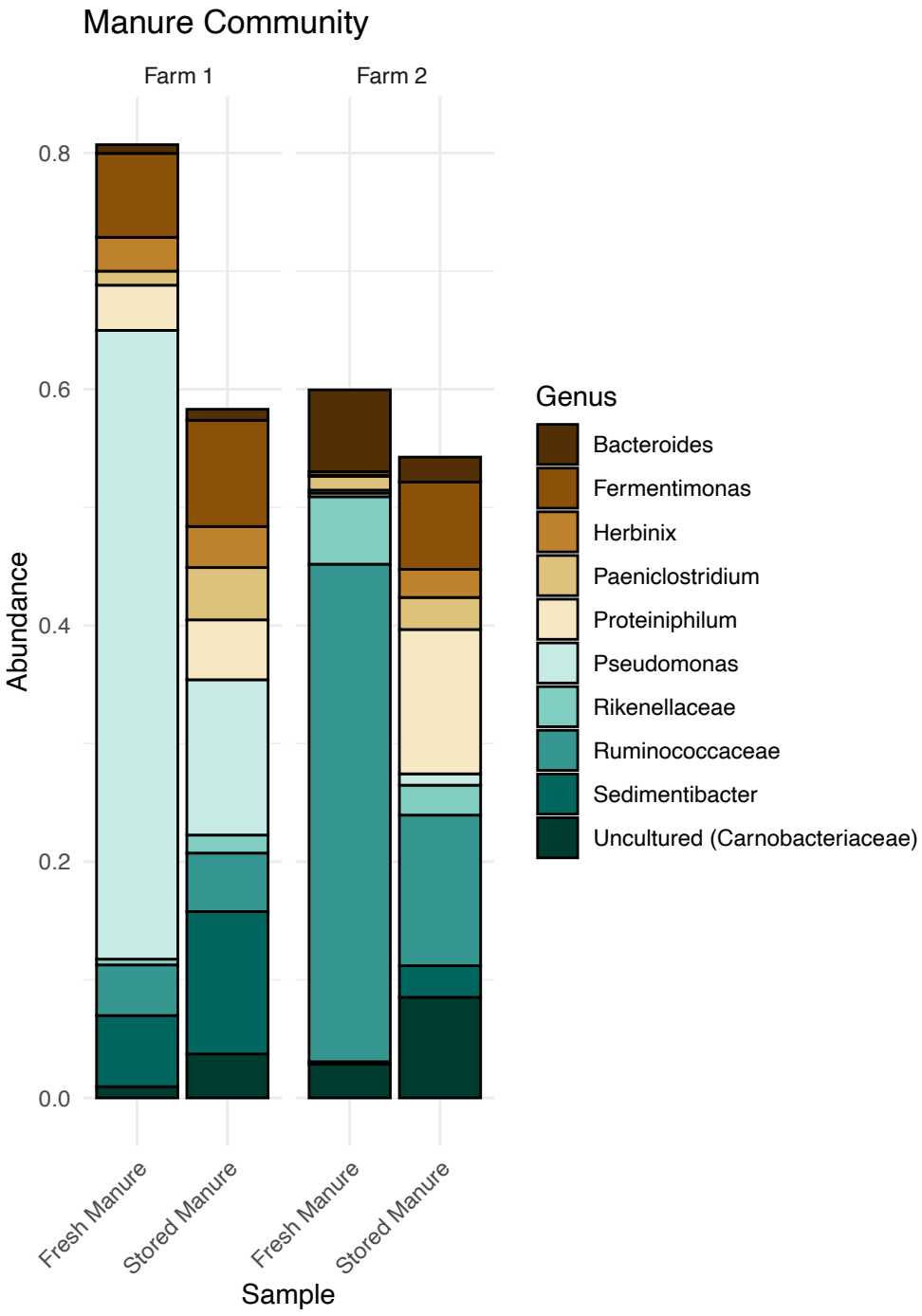

Figure 5. Relative abundance bar plot of the 16S rRNA gene sequences at the genus level in fresh and stored manure in Farm 1 (left) and Farm 2 (right).

The host range of ARGs

ARGs *tetM, strB* and *bla*$_{OXA-58,}$ were detected in both farms and in soil and manure. In total, 664 OTU's were obtained with epicPCR and 879 565 filtered reads were used to analyze the host range of the ARGs (Figures 6, 7 & 8). The number of reads was generally low, and variation was high between the hosts of different ARGs in different samples. Overall, more reads were obtained from Farm 1. Also, more hosts were found from Farm 1 in compared to Farm 2. The host range of *tetM* and *strB* was broader than the host range of *bla*$_{OXA-58}$, which had only 7 different genera (Figure 6, 7 & 8). ARGs *strB* and *tetM* shared 6 host genera at the farms and 3 genera were found to carry all the studied genes.

Due to the low quality of sequences after first sequencing process, samples were sequenced twice. Nevertheless, the second sequencing run did not yield better quality. Both sequencing rounds were analyzed together, but only 1.2% of the total number of reads passed the size and quality filtering. Remaining 98.7% of the reads were unspecific product caused by *bla*$_{OXA-58}$ primer. This primer bound nonspecifically to 16S rRNA gene and thus, did not produce the fused product with both ARG and 16S rRNA gene part. Some number of reads was obtained also from negative controls. Negative control for *tetM* had in total 539 reads, *bla*$_{OXA-58}$ 7 reads and *strB 28* reads.

The number of reads for the hosts of *tetM* varied greatly from 67 to 127 314 between samples. Overall, more reads were obtained for Farm 1 samples, and also more hosts (Figure 6). *tetM* was found in total 48 different hosts (Supplementary Figure S2). Only fresh manure and soil after fertilization had more than 12 hosts. The most common host genera for *tetM* were *Escherichia-Shigella* and *Sedimentibacter,* and these were found in every sample in both farms (Figure 6). *Fibrobacter,* unclassified *Lachnospiraceae* and NK4A136 group *Lachnospiraceae* were found to carry *tetM* almost in all sample types.

Only 7 OTUs were found to be hosts of *bla*$_{OXA-58}$ after filtering. The most common hosts of *bla*$_{OXA-58}$ were *Sphingobacterium* and *Acinetobacter* (Figure 7). *Acinetobacter* was detected more frequently from samples in Farm 1 than in Farm 2. *Escherichia-Shigella*, *Fibrobacter* and *Sedimentibacter* carried both *tetM* and *bla*$_{OXA-58}$. However, *Fibrobacter* and *Sedimentibacter w*ere found to carry *bla*$_{OXA-58}$ only in Farm 1 fresh manure and *Escherichia-Shigella* in Farm 1 soil before fertilization and 2 weeks after fertilization. Again, more reads were obtained in samples from Farm 1 and the read number varied from 7 to approximately 4000 reads.

In total, 31 genera carried streptomycin resistance gene *strB* (Supplementary Figure S3). The gene was carried by *Acinetobacter, Pseudomonas* and *Psychrobacter* in all samples regardless of farm in question (Figure 8). As well as with the other genes, *Escherichia-Shigella, Fibrobacter* and *Sedimentibacter* carried also *strB. Fibrobacter* and *Sedimentibacter* were occasionally found in Farm 1, while *Escherichia-Shigella* was found to be the only host found in Farm 2. More reads were obtained from samples taken from Farm 1 than from Farm 2. The highest number of reads, 340765, was obtained from fresh manure in Farm 1 and lowest (72 reads) in soil 6 weeks after fertilization in Farm 2.

The most abundant OTUs according to the 16S rRNA gene amplicon sequencing were not the same OTUs carrying ARGs. Only part of the OTUs were shared between both datasets (Supplementary Table S1). Due to epicPCR's qualitative feature, results were analyzed using presence/absence approach. When looking at the determined bacterial community, *Sedimentibacter* was detected more in manure (20.9%) than in soils (2.8%). It was detected to carry bla_{OXA-58} and *strB* by epicPCR only in manure samples taken from Farm 1. *Sedimentibacter* that carried *tetM* was detected in every sample with 16S rRNA amplicon sequencing, while *Fibrobacter* that was found to carry all of the studied ARGs was not detected at all. Multiple OTUs part of unclassified Lachnospiraceae family were detected in 16S rRNA gene data, and this family was also found to carry *tetM* in every sample in both farms. *Acinetobacter* carrying bla_{OXA-58} and *strB* was found in the bacterial community more in soil (2.5%) than in manure (0.6%). *Psychrobacter* was the most common carrier of *strB* and found almost as much in the 16S rRNA amplicon sequencing data in manure (5.0%) as in soil (5.5%). *Pseudomonas* was detected by epicPCR as a carrier of *tetM* and *strB* and in the bacterial community in soil 8.8% and in manure 67.6%. Any OTUs belonging to the genus of *Sphingobacterium* were not detected in 16S rRNA amplicon sequences, even though it was found to carry bla_{OXA-58} almost in every sample.

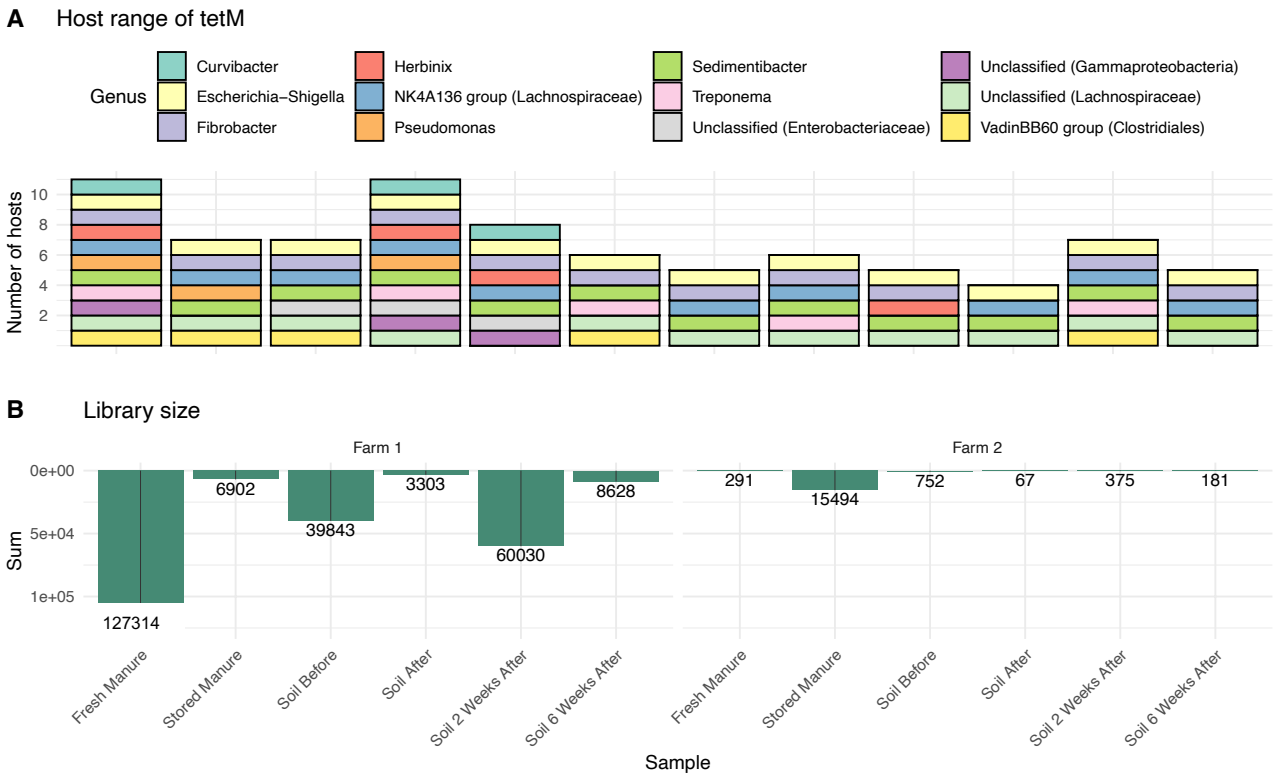

Figure 6. The host range of *tetM* detected by epicPCR. Upper part is showing the number of different bacterial hosts, when the lower part is showing the number of reads obtained in Illumina Miseq sequencing.

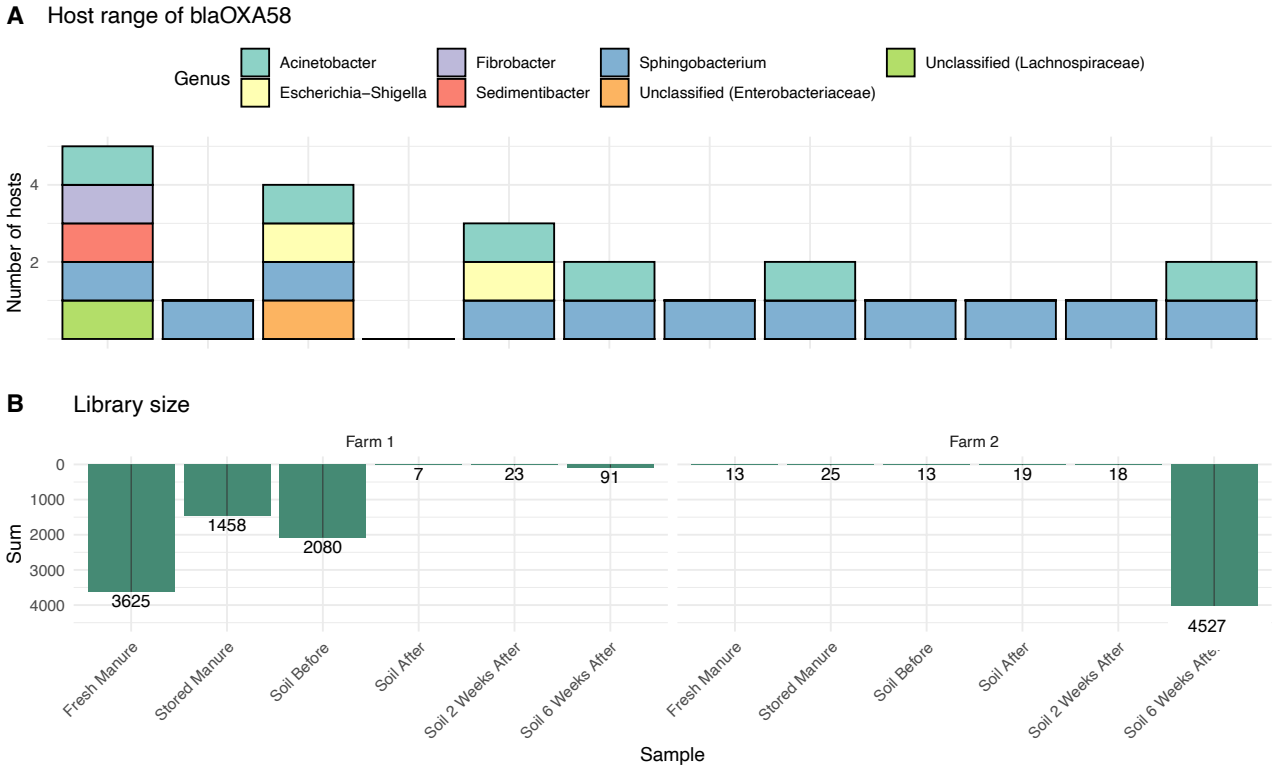

Figure 7. The host range of *bla*$_{OXA-58}$ detected by epicPCR. Upper part is showing the number of different bacterial hosts, when the lower part is showing the number of reads obtained in Illumina Miseq sequencing. Farm 1 is presented on the left and Farm 2 on the right.

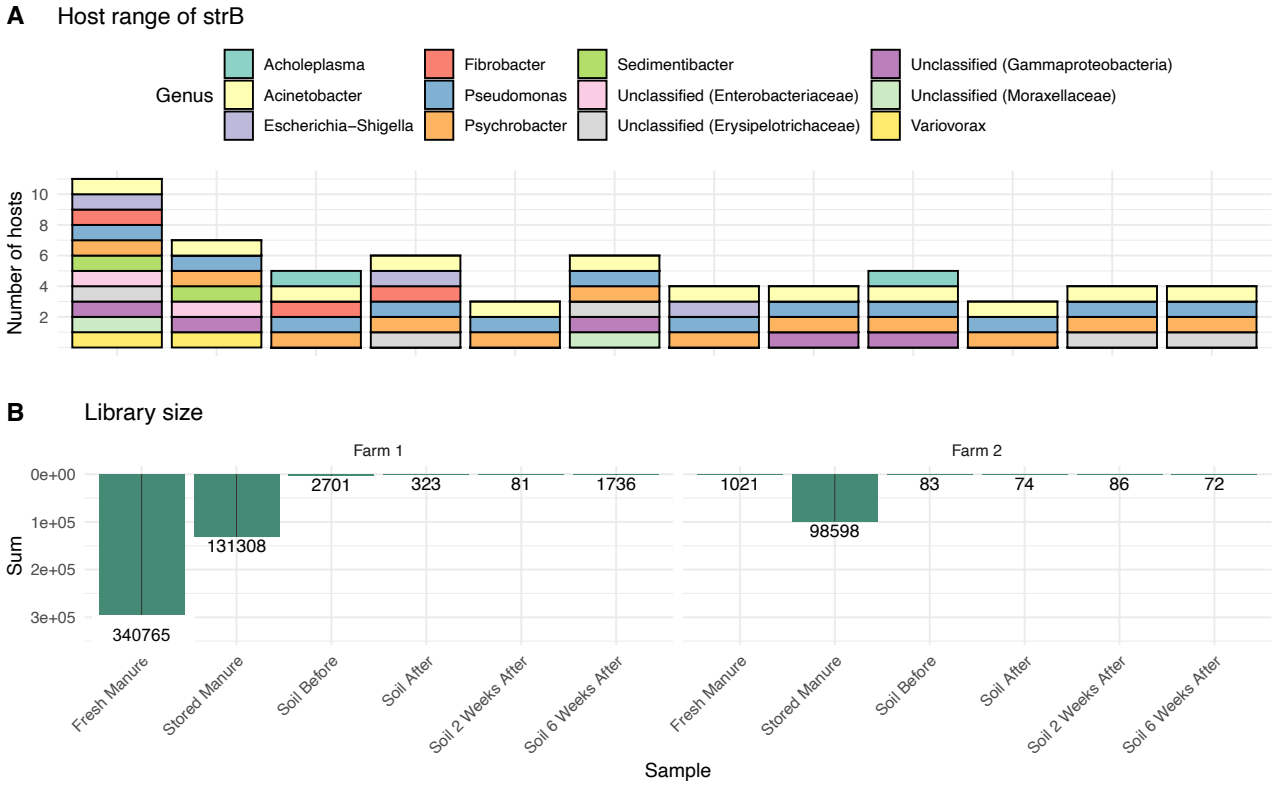

Figure 8. The host range of *strB* detected by epicPCR. Upper part is showing the number of different bacterial hosts, when the lower part is showing the number of reads obtained in Illumina Miseq sequencing.

Discussion

In this study, epicPCR is used for the first time in dairy farms to link ARGs to their host bacteria in soil and manure communities in dairy farms. All the studied genes were carried by multiple genera in manure and manure-fertilized soils in two Finnish dairy farms. The host range was broader for *tetM* and *strB* compared to bla_{OXA-58} and dissimilar between the farms. The epicPCR results were compared to bacterial community compositions obtained with 16S rRNA amplicon sequencing. It was noted that the genera detected in epicPCR, were rarely detected or were rare in the total bacterial community. The qualitative feature of epicPCR could cause this, since the bacteria end up in the beads randomly (Florenza et al., 2019). Also, only those bacteria that carry the target gene are detected. Thus, genera that are relatively rare in the total community can be detected with epicPCR, but not necessarily with 16S rRNA amplicon sequencing, in which only the most abundant genera are detected. Despite the differences, approximately same number of OTUs were found with epicPCR and with 16S rRNA gene analysis.

The host range of ARGs

The bla_{OXA-58} gene was carried by 7 different hosts in both farms. This gene was also detected in the same farms previously by Ruuskanen et al. (2016). The presence of bla_{OXA-58} in the farm environment is interesting, cause carbapenems have not been used in animal production (Davies & Wales, 2019). bla_{OXA-58} gene can thus be originated from soil bacteria and enriched by fertilization, which is reported also by Udikovic-Kolic et al. (2014). The most common carrier of bla_{OXA-58} was *Sphingobacterium*, which are found in various soil and water environments and known to be the producers of chromosomal encoded carbapenemases (Blahová et al., 1997; Henriques et al., 2012). The presence of *Sphingobacterium* can possibly provoke presence of bla_{OXA-58} gene in other genera via HGT. Also, bla_{OXA-58} might have been transferred into cattle gut via harvested feed, and further transferred horizontally to other hosts via HGT. *Acinetobacter* was detected with bla_{OXA-58} in the present study both in soil and manure samples. *Acinetobacter spp*. has been widely reported to harbor bla_{OXA-58} in broad range plasmids with carbapenemase genes (Bertini et al., 2007, Leski et al., 2013). Some of these plasmids are known to be able to spread in a broad host range of bacterial species (EFSA, 2013), which accelerates the spread of ARGs.

On the other hand, *bla*$_{OXA-58}$ has also been reported to be silent and expressed in the presence of insertion sequence element (Bertini et al., 2007). Thus, selection caused by antibiotics, heavy metals and other pollutants could drive additional insertion of IS element and that way activate the *bla*$_{OXA-58}$ gene (Zhou, 2015). In addition of *bla*$_{OXA-58}$ gene, *Acinetobacter* were also carriers of *strB* gene in every sample. Supporting results have been documented by Fournier et al. (2006) after sequencing genomic island AbaR1. This particular island was reported to be part of broad host-range MGEs and originating from *Pseudomonas, Salmonella* and *Escherichia spp*. suggesting, that *Acinetobacter* could have acquired *strB* and *bla*$_{OXA-58}$ in broad host-range plasmids via HGT.

Here, *Psychrobacter* and *Pseudomonas* were carrying of *strB* in every sample in both farms, while *Escherichia-Shigella* were carriers only in few samples. *Psychrobacter* can be considered to be the instinct carrier of *strB* gene, since the genera has been documented to carry a mosaic structured plasmid containing *strB* gene in ancient permafrost (Petrova et al., 2009). The genera have also been found from agricultural soils during wintertime (Udikovic-Kolic et al., 2014). Also, *Pseudomonas spp.* and *E. coli* have been previously reported with a *strB* gene (Sunde & Sorum, 1999; Sundin, 2002). Since the presence of *strB* indicates the presence of integrons (Chiou & Jones, 1993) and is highly linked to transmissible plasmids and gene cassettes, its potentially transferable to the pathogens and to important zoonotic bacteria.

Escherichia-Shigella, Fibrobacter and *Sedimentibacter* were found to be one of the most common carriers of *tetM. Sedimentibacter* was associated with *tetM* for the first time, indicating that *tetM* was transferred via HGT from gut microbes to *Sedimentibacter* in soil as a result of manure application. The fact that *Sedimentibacter* is likely originating from sediment (Breitenstein et al., 2002) supports the theory of horizontally acquired *tetM*. *Escherichia-Shigella* was detected with *tetM* in all samples. Some bacteriophages have been reported to induce the dissemination of *tetM* in *E. coli* and *Shigella sonnei* in natural populations (Gabasvhili et al., 2020). Thus, phages could be potentially spreading *tetM* between bacteria in the farm environment as well. *Fibrobacter* was associated with *tetM* for the first time and found in all manure and soil samples. *Fibrobacter* is known as a cellulolytic bacterium found in rumen (Koike & Kobayashi, 2001). Unclassified *Lachnospiraceae* and NK4A136 group *Lachnospiraceae* families*,* also known to be common among the rumen bacteria (Seshadri et al., 2018) and reported to carry *tetM* in wastewater influent (Hultman et al., 2018), were found to carry *tetM* by multiple genera in both soil and manure*.* Intestinal microbiota of antimicrobial-medicated animals might have selective conditions for *tetM*,

since the gene is mainly found from gut microbes and is known to be located in conjugative MGEs (Chopra et al., 2001). Manure application has shown to increase the abundance of ARGs in soil (Muurinen et al., 2017; McKinney et al., 2018) and higher application rate to produce higher level or AGRs (McKinney et al., 2018). Also, the abundance of *tetM* in soil is considered as a sign of a manure treatment (Wu et al. 2010). Nevertheless, the abundance of *tetM* in manure-fertilized soils has been found to decrease or even disappear over time (Alexander et al., 2011; Muurinen et al., 2017).

Challenges and limits of epicPCR

In this study, only 879 565 reads (less than 2% of the total reads) were obtained from epicPCR and used for analysis. In this thesis, epicPCR was used for the first time in soil and manure. Soil and manure were observed to be challenging sample matrixes and potentially affecting the number of reads. Soils can contain multiple substances i.e. organic compounds that can inhibit the PCR reaction, while complex polysaccharides, bile salts, lipids and urate are the most inhibiting substances in manure (Schrader et al., 2012). Some soil and manure particles attached with free DNA might have ended up with the extracted cells in the beads disrupting epicPCR and producing unspecific products. A lot of unspecific products were seen for each gene already on gel electrophoresis, which indicated that problems would arise in sequencing. Thus, inspection of PCR products in gel electrophoresis is essential, since environmental DNA is known as a challenging starting material and lack of sensitivity (van Elsas & Wolter, 1995). In addition to optimization of pretreatment of these types of samples, the primers should be tested and designed before performing this method for valuable samples. Samples used in this study were stored in glycerol and thus could be thawed only once, leaving only one attempt for success.

Pooling of the different PCR reactions before sequencing caused the domination of bla_{OXA-58} primer in every sample. This way the bla_{OXA-58} primer affected also to the *tetM* and *strB* epicPCR products, disrupted the sequencing and yielded low number of reads for each sample type. Thus, replicates of same sample for different genes should not be pooled before sequencing, to ensure successful sequencing process and minimize the risks for errors.

Primers for bla_{OXA-58} and *tetM* in epicPCR has been used successfully by Hultman et al. (2018) in wastewaters. The incidence that primers for bla_{OXA-58} did not work well in this work might have been caused by polymorphism of *bla* genes in soil and manure. Soil bacteria are

suggested to be intrinsically resistant to a broad spectrum of beta-lactam antibiotics and soils most likely harbor a diverse variety of *bla* genes carried by various bacteria (Demanèche et al., 2008). The single nucleotide polymorphism of bla_{oxa} genes can occur in bacteria living in different environments and habitats, which would explain why bla_{OXA-58} primers worked well for bacteria in wastewaters but not for bacteria in soil and manure. Also, distribution of microbial genes has been explained more by ecology than geographic location (Fondi et al., 2016), indicating that homologs of bla_{OXA-58} gene could be dissimilar between bacteria in different ecological niches. This supports the hypothesis that even same species in different environments might have differences in same sequence coding the same feature.

Also, few reads were obtained from negative controls. This is most probably due to multiple PCR rounds, increasing the possibility for aerosol contamination or cross contamination between samples. Negative controls were created only in the last PCR step (nested PCR), which gives information of the contamination only in that particular PCR. Performing negative controls already from empty beads would create the most representative control for the whole protocol. However, probably the most contaminating step is still nested PCR in which the starting material is purified PCR product. This problem could be tackled by performing different steps of experiment in specific rooms and this way prevent the aerosolization of PCR product. In this work, Decontam package and R was used to solve the problem with negative controls statistically. Threshold for each gene was considered separately, since the library sizes varied between negative controls, samples and genes. Here higher threshold was used for *tetM* gene. This way the most reliable number of OTUs could be obtained for each gene.

In this study the 12 most abundant host bacteria for *tetM,* bla_{OXA-58} and *strB* were presented. However, some other approaches for data visualization should be developed. EpicPCR is not a quantitative method and presenting nonquantitative data creates challenges, as many of the programs are designed for drawing plots for quantitative data only. Successful sequencing of epicPCR products can possibly yield hundreds of host genera for one ARG, which increases the challenges in visualization. Another limiting element in epicPCR is potentially the size of polyacrylamide bead. The size of the bead affects the amount of PCR reagents diffused inside further influencing the concentration of the end product. Also, number of cells in the beads is estimated by visualizing the

fluorescent dyed cells by microscope in only one small fraction of the total volume. This probably has an effect on how reliable results are and will be discussed shortly.

Previously mentioned pretreatment including extraction of cells based on nonionic gradation and filtering the sample matrix through a cell strainer might cause loss of some bacteria. Sonication step might lyze cells with weaker cell wall influencing selectively results (Ramsay, 1984). On the other hand, as a sample types soil and manure contained plenty tight cell clusters hard to separate, which were still seen in the beads and could yield false positive results. Nonetheless, since the most dominant lifestyle of bacteria is in biofilms (Flemming & Wingender, 2010), bacteria could also be considered as multicellular organisms. Therefore, bacterial cell clusters ending up in one bead would offer important information on the presence of a gene even though the cell cluster would contain more than one species. Thus, epicPCR could be used to study the hosts of ARGs despite the beads would contain small clusters. Interaction between bacteria in biofilms is most likely high and the presence of ARGs plays a role for the whole community in biofilm (Balcázar et al., 2015). Hence, so-called false positive results (more than one cell in the bead and not all cells carry the target ARG) wouldn't be a false positive in that sense. Nevertheless, here an OTU carrying ARG was considered as a reliable result only if it was present in 2/3 replicates. This filtering was used due to the errors in the sequencing process. Like in the study by Hultman et al. (2018), filtering could be increased to 3/3 to lower the possibility for false positive results caused by multiple different cells entering one bead.

In epicPCR single cells are used as a starting point. One gram of soil contains thousands of individual microbial taxa (Fierer, 2017). Only some bacteria detected in epicPCR were present the total bacterial community. Therefore, the detection limit of epicPCR is quite high and if this study would be repeated with the same samples, it would be unlikely to obtain the same results. This is caused by the randomness of what bacteria will be captured in the beads (Florenza et al., 2019). Also, soil is known to be very heterogenous environment in which even small distances can have totally different microbial abundance and community composition (Fierer, 2017). This again increases the randomness of what is detected. On the other hand, here, 1 kg of soil was used as a starting material for one sample, homogenized by sieving, mixed, stored in glycerol and mixed before pretreatment with PBS. Thus, the sample was more homogenized and representative than just a small amount of "dry" soil. However, it is still important to determine with 16S rRNA gene sequencing if the

total bacterial community to support the findings in epicPCR, as was done in this study. However, here only 1 replicate for each sample was sequenced, which hindered more sophisticated statistical analysis. To obtain results that could be analyzed statistically, minimum 3 replicates should be included per each sample. Statistical methods that are possible to conduct with only one sample are limited to ordinations and non-parametric tests, such as PERMANOVA, and also in these, three replicates instead of one would decrease the random variance and yield more veracious results.

Conclusions

Typical culture-based approaches leave significant knowledge gaps on the diversity of bacteria harboring ARGs. Especially in soil, the majority of the microbial diversity is still uncharacterized (Fierer, 2017). EpicPCR is an excellent culture independent method to link a gene of interest to its host bacteria. Compared to metagenomic approach, the detection limit of epicPCR is improved since small quantities of ARGs can potentially be undetected by metagenomic sequencing (Bengtsson-Palme et al., 2017). This study provides valuable information in order to further improve the comparatively new method to study ARGs in soil and manure. For future perspectives, the genetic environment of ARGs could be determined, e.g. by using inverse-PCR (Ocham et al., 1989; Pärnänen et al., 2016) to investigate the mobility of ARGs. Together with comparison of bacteriophage genomes and genomic islands of ARGs in bacterial genomes would help us to understand the mechanisms and rate of HGT. The combination of resulting information on the genetic context of the ARGs together with determined host bacteria by epicPCR, would help us to profile the key elements and genera spreading antimicrobial resistance.

Acknowledgements

I wish to acknowledge the Academy of Finland and Maj & Tor Nessling foundation, that were originally funding the costs used in this study and CSC—IT Center for Science, Finland, for providing the computational resources used for the data analysis. Many thanks for Aino Rutanen for designing *strB* primer set and enabling this study. I would like to express my deepest gratitude for my supervisors Dr. Johanna Muurinen, Katariina Pärnänen and Dr. Jenni Hultman for their precious guidance, feedback and patience, whenever needed, that

helped me to accomplish this work. I would also like to thank my PI Prof. Marko Virta, who gave me the opportunity to carry out this project and provided me with his valuable advices during the project. I would also like to express my deepest gratitude to my fellow lab mates that have supported me during the journey. Thanks for all your support, I have learned so much from all of you. Finally, thanks for my friends and family, giving me constant support during my thesis.

References

Alexander, T. W., Yanke, J. L., Reuter, T., Topp, E., Read, R. R., Selinger, B. L., & McAllister, T. A. (2011). Longitudinal characterization of antimicrobial resistance genes in feces shed from cattle fed different subtherapeutic antibiotics. *BMC Microbiology*, *11*(1), 19.

Allen, H. K., Donato, J., Wang, H. H., Cloud-Hansen, K. A., Davies, J., & Handelsman, J. (2010). Call of the wild: antibiotic resistance genes in natural environments. *Nature Reviews Microbiology*, *8*(4), 251-259.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403-410.

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data.

Balcázar, J. L., Subirats, J., & Borrego, C. M. (2015). The role of biofilms as environmental reservoirs of antibiotic resistance. *Frontiers in Microbiology*, *6*, 1216.

Baquero, F., Martínez, J. L., & Cantón, R. (2008). Antibiotics and antibiotic resistance in water environments. *Current Opinion in Biotechnology*, *19*(3), 260-265.

Bengtsson-Palme, J., Larsson, D. J., & Kristiansson, E. (2017). Using metagenomics to investigate human and environmental resistomes. *Journal of Antimicrobial Chemotherapy*, *72*(10), 2690-2703.

Blahová, J., Králiková, K., Krčméry, V., & Kuboňová, K. (1997). Hydrolysis of imipenem, meropenem, ceftazidime, and cefepime by multiresistant nosocomial strains of *Sphingobacterium multivorum*. *European Journal of Clinical Microbiology and Infectious Diseases*, *16*(2), 178-180.

Breitenstein, A., Wiegel, J., Haertig, C., Weiss, N., Andreesen, J. R., & Lechner, U. (2002). Reclassification of *Clostridium hydroxybenzoicum* as *Sedimentibacter hydroxybenzoicus* gen. nov., comb. nov., and description of *Sedimentibacter saalensis* sp. nov.. *International Journal of Systematic and Evolutionary Microbiology*, *52*(3), 801-807.

Chee-Sanford, J. C., Mackie, R. I., Koike, S., Krapac, I. G., Lin, Y. F., Yannarell, A. C., Maxwell, S., & Aminov, R. I. (2009). Fate and transport of antibiotic residues and antibiotic resistance genes following land application of manure waste. *Journal of Environmental Quality*, *38*(3), 1086-1108.

Chiou, C. S., and Jones, A. L. (1993). Nucleotide sequence analysis of a transposon (Tn*5393*) carrying streptomycin resistance genes in *Erwinia amylovora* and other gram-negative bacteria. *Journal of Bacteriology.* 175, 732–740. doi: 10.1128/jb.175.3.732-740.1993

Chopra, I., & Roberts, M. (2001). Tetracycline antibiotics: mode of action, applications, molecular biology, and epidemiology of bacterial resistance. *Microbiology and Molecular Biology Reviews.*, *65*(2), 232-260.

Cole, D., Drum, D. J., Stalknecht, D. E., White, D. G., Lee, M. D., Ayers, S., Sobsey, M., & Maurer, J. J. (2005). Free-living Canada geese and antimicrobial resistance. *Emerging Infectious Diseases*, *11*(6), 935-938.

Davis, N. M., Proctor, D. M., Holmes, S. P., Relman, D. A., & Callahan, B. J. (2018). Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome*, *6*(1), 226.

Davies, R., & Wales, A. (2019). Antimicrobial resistance on farms: a review including biosecurity and the potential role of disinfectants in resistance selection. *Comprehensive Reviews in Food Science and Food Safety*, *18*(3), 753-774.

Demanèche, S., Sanguin, H., Poté, J., Navarro, E., Bernillon, D., Mavingui, P., Wildi, W., Vogel, T. M., & Simonet, P. (2008). Antibiotic-resistant soil bacteria in transgenic plant fields. *Proceedings of the National Academy of Sciences*, *105*(10), 3957-3962.

Dolejska, M., Cizek, A., & Literak, I. (2007). High prevalence of antimicrobial-resistant genes and integrons in *Escherichia coli* isolates from black-headed gulls in the Czech Republic. *Journal of Applied Microbiology*, *103*(1), 11-19.

EFSA Panel on Biological Hazards (BIOHAZ). (2013). Scientific Opinion on Carbapenem resistance in food animal ecosystems. *EFSA Journal*, *11*(12), 3501.

European Medicines Agency. (2018). Sales of veterinary antimicrobial agents in 30 European countries in 2016. Trends from 2010 to 2016. *8th ESVAC report*.

FAO Food and Agriculture Organization of United Nations (2019). Antimicrobial resistance, Animal production. Available at: http://www.fao.org/antimicrobial-resistance/key-sectors/animal-production/en/ . Accessed 8 Dec 2019.

Fierer, N. (2017). Embracing the unknown: disentangling the complexities of the soil microbiome. *Nature Reviews Microbiology*, *15*(10), 579.

Finlex 2000. Valtioneuvoston asetus maataloudesta peräisin olevien nitraattien vesiin pääsyn rajoittamisesta 931/2000. Available at: https://www.finlex.fi/fi/laki/alkup/2000/20000931 . Accessed 10 Dec 2019.

Flemming, H. C., & Wingender, J. (2010). The biofilm matrix. *Nature Reviews Microbiology*, *8*(9), 623-633.

Florenza, J., Tamminen, M., & Bertilsson, S. (2019). Uncovering microbial inter-domain interactions in complex communities. *Philosophical Transactions of the Royal Society B*, *374*(1786), 20190087.

Fondi, M., Karkman, A., Tamminen, M. V., Bosi, E., Virta, M., Fani, R., Alm, E., & McInerney, J. O. (2016). “Every gene is everywhere but the environment selects”: global geolocalization of gene sharing in environmental samples through network analysis. *Genome Biology and Evolution*, *8*(5), 1388-1400.

Fournier, P. E., Vallenet, D., Barbe, V., Audic, S., Ogata, H., Poirel, L., Richet, H., Robert, C., Mangenot, S., Abergel, C., Nordmann, P., Weissenbach, J., Raoult, D. & Claverie, J-M. (2006). Comparative genomics of multidrug resistance in *Acinetobacter baumannii*. *PLOS Genetics*, *2*(1).

Furuya, E. Y., & Lowy, F. D. (2006). Antimicrobial-resistant bacteria in the community setting. *Nature Reviews Microbiology*, *4*(1), 36-45.

Gabashvili, E., Osepashvili, M., Koulouris, S., Ujmajuridze, L., Tskhitishvili, Z., & Kotetishvili, M. (2020). Phage transduction is involved in the intergeneric spread of antibiotic resistance-associated *bla*$_{CTX-M}$, *mel*, and *tetM* loci in natural populations of some human and animal bacterial pathogens. *Current Microbiology*, *77*(2), 185-193.

Gillings, M. R., & Stokes, H. W. (2012). Are humans increasing bacterial evolvability?. *Trends in Ecology & Evolution*, *27*(6), 346-352.

Graham, J. P., Evans, S. L., Price, L. B., & Silbergeld, E. K. (2009). Fate of antimicrobial-resistant *enterococci* and *staphylococci* and resistance determinants in stored poultry litter. *Environmental Research*, *109*(6), 682-689.

Henriques, I. S., Araújo, S., Azevedo, J. S., Alves, M. S., Chouchani, C., Pereira, A., & Correia, A. (2012). Prevalence and diversity of carbapenem-resistant bacteria in untreated drinking water in Portugal. *Microbial Drug Resistance*, *18*(5), 531-537.

Herlemann, D. P., Labrenz, M., Jürgens, K., Bertilsson, S., Waniek, J. J., & Andersson, A. F. (2011). Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. *The ISME Journal*, *5*(10), 1571-1579.

Hultman, J., Tamminen, M., Pärnänen, K., Cairns, J., Karkman, A., & Virta, M. (2018). Host range of antibiotic resistance genes in wastewater treatment plant influent and effluent. *FEMS Microbiology Ecology*, *94*(4), fiy038.

Karkman, A., Johnson, T. A., Lyra, C., Stedtfeld, R. D., Tamminen, M., Tiedje, J. M., & Virta, M. (2016). High-throughput quantification of antibiotic resistance genes from an urban wastewater treatment plant. *FEMS Microbiology Ecology*, *92*(3).

Knapp, C. W., Dolfing, J., Ehlert, P. A., & Graham, D. W. (2010). Evidence of increasing antibiotic resistance gene abundances in archived soils since 1940. *Environmental Science & Technology*, *44*(2), 580-587.

Koike, S., & Kobayashi, Y. (2001). Development and use of competitive PCR assays for the rumen cellulolytic bacteria: *Fibrobacter succinogenes, Ruminococcus albus* and *Ruminococcus flavefaciens*. *FEMS Microbiology Letters*, *204*(2), 361-366.

Lane, D.J. (1991) 16S/23S rRNA sequencing. In: Stackebrandt, E. and Goodfellow, M., Eds., *Nucleic Acid Techniques in Bacterial Systematic*, John Wiley and Sons, New York, 115-175.

Leski, T. A., Bangura, U., Jimmy, D. H., Ansumana, R., Lizewski, S. E., Li, R. W., Stenger, D. A., Taitt, C. R., & Vora, G. J. (2013). Identification of bla_{OXA-51}*-like,* bla_{OXA-58}*,* bla_{DIM-1}, and bla_{VIM} carbapenemase genes in hospital *Enterobacteriaceae* isolates from Sierra Leone. *Journal of Clinical Microbiology*, *51*(7), 2435-2438.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet Journal*, *17*(1), 10-12.

Martínez, J. L. (2008). Antibiotics and antibiotic resistance genes in natural environments. *Science*, *321*(5887), 365-367.

McKinney, C. W., Dungan, R. S., Moore, A., & Leytem, A. B. (2018). Occurrence and abundance of antibiotic resistance genes in agricultural soil receiving dairy manure. *FEMS Microbiology Ecology*, *94*(3), fiy010.

McMurdie, P. J., & Holmes, S. (2013). Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLOS One*, *8*(4).

Muurinen, J., Stedtfeld, R., Karkman, A., Parnanen, K., Tiedje, J., & Virta, M. (2017). Influence of manure application on the environmental resistome under Finnish agricultural practice with restricted antibiotic use. *Environmental Science & Technology*, *51*(11), 5989-5999.

Ochman, H., Ajioka, J. W., Garza, D., & Hartl, D. L. (1989). Inverse polymerase chain reaction. In *PCR Technology* (pp. 105-111). Palgrave Macmillan, London.

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., O'hara, R. B., Simpson, G. L., Solymos, P. M., Stevens, H. H., Szoecs, E., & Wagner, H. (2010). Vegan: community ecology package. R package version 1.17-4. https://CRAN.R-pro ject.org/package=vegan.

One Health Global Network (2012). One Health: a concept that became an approach and then a movement. Available at: http://www.onehealthglobal.net/what-is-one-health/ . Accessed 10 Jan 2020.

Petrova, M., Gorlenko, Z., & Mindlin, S. (2009). Molecular structure and translocation of a multiple antibiotic resistance region of a *Psychrobacter psychrophilus* permafrost strain. *FEMS Microbiology Letters*, *296*(2), 190-197.

Poirel, L., & Nordmann, P. (2006). Genetic structures at the origin of acquisition and expression of the carbapenem-hydrolyzing oxacillinase gene blaOXA-58 in *Acinetobacter baumannii*. *Antimicrobial Agents and Chemotherapy*, *50*(4), 1442-1448.

Price, L. B., Johnson, E., Vailes, R., & Silbergeld, E. (2005). Fluoroquinolone-resistant *Campylobacter* isolates from conventional and antibiotic-free chicken products. *Environmental Health Perspectives*, *113*(5), 557-560.

Pärnänen, K., Karkman, A., Tamminen, M., Lyra, C., Hultman, J., Paulin, L., & Virta, M. (2016). Evaluating the mobility potential of antibiotic resistance genes in environmental resistomes without metagenomics. *Scientific Reports*, *6*, 35790.

Qin, H., Wang, S., Feng, K., He, Z., Virta, M. P., Hou, W., Dong, H., & Deng, Y. (2019). Unraveling the diversity of sedimentary sulfate-reducing prokaryotes (SRP) across Tibetan saline lakes using epicPCR. *Microbiome*, *7*(1), 71.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., & Glöckner, F. O. (2012). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research*, *41*(D1), D590-D596.

Ramsay, A. J. (1984). Extraction of bacteria from soil: efficiency of shaking or ultrasonication as indicated by direct counts and autoradiography. *Soil Biology and Biochemistry*, *16*(5), 475-481.

Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahé, F. (2016). VSEARCH: a versatile open source tool for metagenomics. *PeerJ*, *4*, e2584.

R Core Team. (2017). R: A language and environment for statistical computing. URL https://www.R-project.org/.

RStudio Team (2015). RStudio: Integrated development for R. RStudio, Inc., Boston, MA URL http://www.rstudio.com/.

Ruuskanen, M., Muurinen, J., Meierjohan, A., Pärnänen, K., Tamminen, M., Lyra, C., Kronberg, L., & Virta, M. (2016). Fertilizing with animal manure disseminates antibiotic resistance genes to the farm environment. *Journal of Environmental Quality*, *45*(2), 488-493.

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., Lesniewski, R. A., Oakley, B. B., Parks, D. H., Robinson, C. J., Sahl, J. W., Stres, B., Thallinger, G. G., van Horn D. J., & Weber, C. F. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied Environmental Microbiology*, *75*(23), 7537-7541.

Schmieder, R., & Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. *Bioinformatics*, *27*(6), 863-864.

Schrader, C., Schielke, A., Ellerbroek, L., & Johne, R. (2012). PCR inhibitors–occurrence, properties and removal. *Journal of Applied Microbiology*, *113*(5), 1014-1026.

Seshadri, R., Leahy, S. C., Attwood, G. T., Teh, K. H., Lambie, S. C., Cookson, A. L., Eloe-Fadrosh, E. A., Pavlopoulos, G. A., Hadjithomas, M., Varghese, N. J., Paez-Espino, D., Hungate1000 project collabolators, Perry, R., Henderson, G., Creevey, C. J., Terrapon, N., Lapebie, P., Drula, E., Lombard, V., Rubin, E., Kyrpides, N. C., Henrissat, B., Woyke, T., Ivanova, N. N., & Kelly, W. J. (2018). Cultivation and sequencing of rumen microbiome members from the Hungate1000 Collection. *Nature Biotechnology*, *36*(4), 359.

Smith, T. C., Gebreyes, W. A., Abley, M. J., Harper, A. L., Forshey, B. M., Male, M. J., Martin, H. W., Molla, B. Z., Streevatsan, S., Thakur, S., Thiruvengadam, M., & Davies, P. R. (2013). Methicillin-resistant *Staphylococcus aureus* in pigs and farm workers on conventional and antibiotic-free swine farms in the USA. *PLOS One*, *8*(5), e63704.

Spencer, S. J., Tamminen, M. V., Preheim, S. P., Guo, M. T., Briggs, A. W., Brito, I. L., Weitz, A. D., Pitkänen, L. K., Vigneault, F., Virta, J. M. P., & Alm, E. J. (2016). Massively parallel sequencing of single cells by epicPCR links functional genes with phylogenetic markers. *The ISME Journal*, *10*(2), 427-436.

Sunde, M., & Sorum, H. (1999). Characterization of integrons in *Escherichia coli* of the normal intestinal flora of swine. *Microbial Drug Resistance*, *5*(4), 279-287.

Sundin, G. W., & Bender, C. L. (1996). Dissemination of the *strA-strB* streptomycin-resistance genes among commensal and pathogenic bacteria from humans, animals, and plants. *Molecular Ecology*, *5*(1), 133-143.

Sundin, G. W. (2002). Distinct recent lineages of the *strA-strB* streptomycin-resistance genes in clinical and environmental bacteria. *Current Microbiology*, *45*(1), 63-69.

Surette, M. D., & Wright, G. D. (2017). Lessons from the environmental antibiotic resistome. *Annual Review of Microbiology*, *71*, 309-329.

Tamminen, M., Virta, M., Fani, R., & Fondi, M. (2012). Large-scale analysis of plasmid relationships through gene-sharing networks. *Molecular Biology and Evolution*, *29*(4), 1225-1240.

Udikovic-Kolic, N., Wichmann, F., Broderick, N. A., & Handelsman, J. (2014). Bloom of resident antibiotic-resistant bacteria in soil following manure fertilization. *Proceedings of the National Academy of Sciences*, *111*(42), 15202-15207.

van Boeckel, T. P., Brower, C., Gilbert, M., Grenfell, B. T., Levin, S. A., Robinson, T. P., Teillant, A., & Laxminarayan, R. (2015). Global trends in antimicrobial use in food animals. *Proceedings of the National Academy of Sciences*, *112*(18), 5649-5654.

van Elsas, J. D., & Wolters, A. (1995). Polymerase chain reaction (PCR) analysis of soil microbial DNA. In *Molecular Microbial Ecology Manual* (pp. 235-244). Springer, Dordrecht.

Wellington, E. M., Boxall, A. B., Cross, P., Feil, E. J., Gaze, W. H., Hawkey, P. M., Johnson-Rollings, A. S., Jones, D. L., Lee, N. M., Otten, W., Thomas, C. M., & Williams A. P. (2013). The role of the natural environment in the emergence of antibiotic resistance in Gram-negative bacteria. *The Lancet Infectious Diseases*, *13*(2), 155-165.

WHO World Health Organization (2019). Ten threats to global health in 2019. Available at: https://www.who.int/emergencies/ten-threats-to-global-health-in-2019 . Accessed 10 Dec 2019.

Wickham, H. (2009) ggplot2: elegant graphics for data analysis Springer-Verlag New York; 2009. https://ggplot2.tidyverse.org.

Witte, W. (2000). Ecological impact of antibiotic use in animals on different complex microflora: environment. *International Journal of Antimicrobial Agents*, *14*(4), 321-325.

Wright, G. D. (2010). Antibiotic resistance in the environment: a link to the clinic?. *Current Opinion in Microbiology*, *13*(5), 589-594.

Wu, N., Qiao, M., Zhang, B., Cheng, W. D., & Zhu, Y. G. (2010). Abundance and diversity of tetracycline resistance genes in soils adjacent to representative swine feedlots in China. *Environmental Science & Technology*, *44*(18), 6933-6939.

Zhang, J., Kobert, K., Flouri, T., & Stamatakis, A. (2014). PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics*, *30*(5), 614-620.

Zhou, S., Chen, X., Meng, X., Zhang, G., Wang, J., Zhou, D., & Guo, X. (2015). “Roar” of *bla*$_{NDM-1}$ and “silence” of *bla*$_{OXA-58}$ co-exist in *Acinetobacter pittii. Scientific Reports*, *5*, 8976.

Supplementary materials

Table S1. The presence of genera carrying ARGs in the total bacterial community

Genera detected in epic	Genera present in the total microbial community	Carrier of *tetM*	Carrier of *bla*$_{OXA-58}$	Carrier of *strB*
Curvibacter		x		
Escherichia-Shigella		x	x	x
Fibrobacter		x	x	x
Herbinix	x	x		
NK4A136 group (Lachnospriaceae)		x		
Pseudomonas	x	x		x
Sedimentibacter	x	x	x	x
Treponema	x	x		
Unclassified (Enterobacteriaceae)		x	x	x
Unclassified (Gamma proteobacteriaceae)	x	x		x
Unclassified (Lachnospriaceae)	x	x	x	
VadinBB60 group (Clostridiales)	x	x		
Acinetobacter	x		x	x
Sphingobacterium			x	
Acholeplasma	x			x
Unclassified (Erysipelotrichaceae)	x			x
Unclassified (Moraxellaceae)				x
Variovorax				x
Psychrobacter	x			x

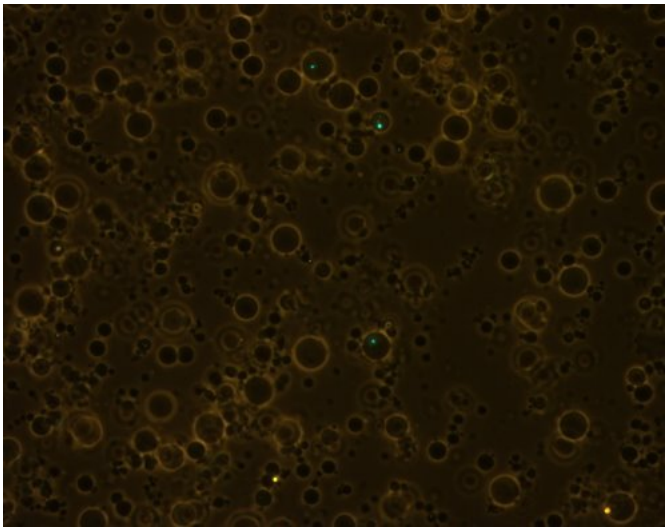

Figure S1. Fluorescent dyed cells in polyacrylamide beads

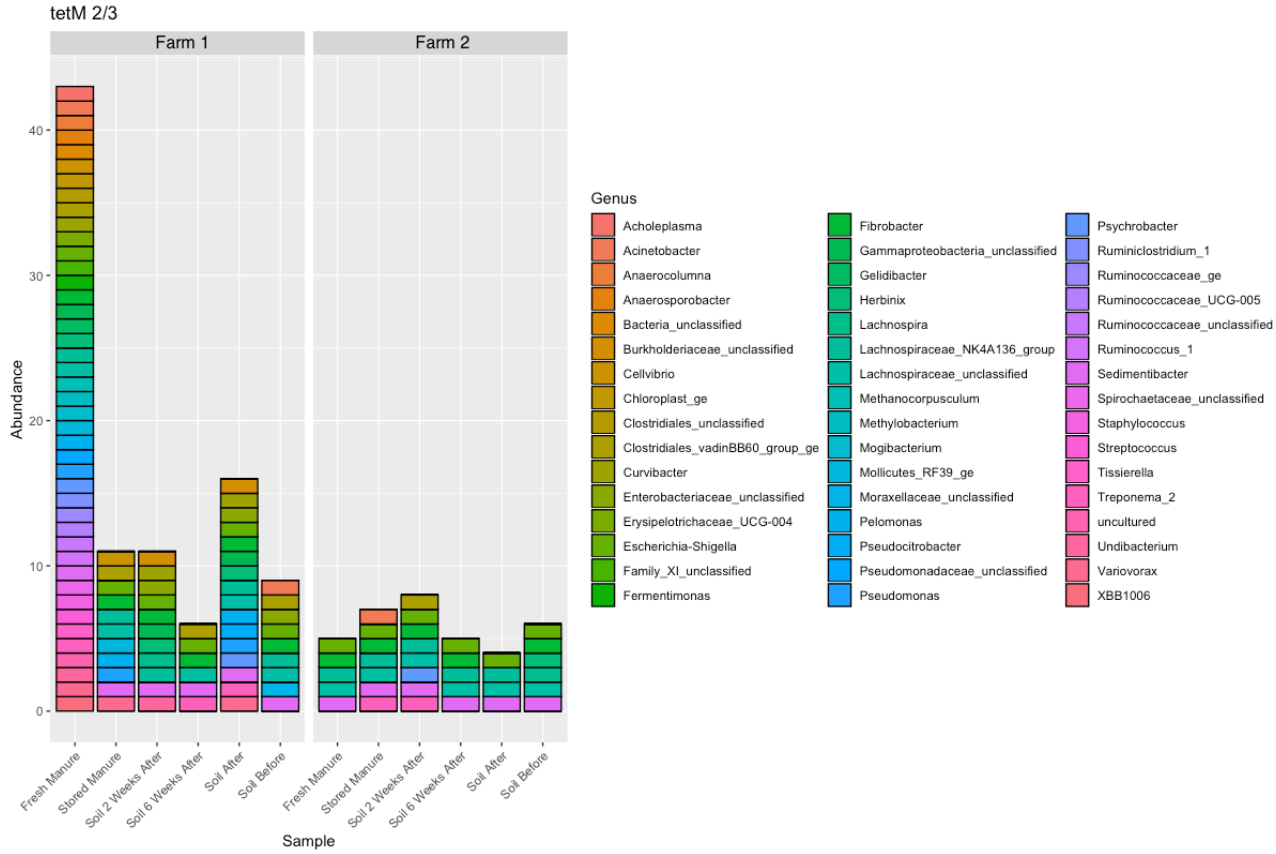

Figure S2. All detected OTUs present in *tetM*

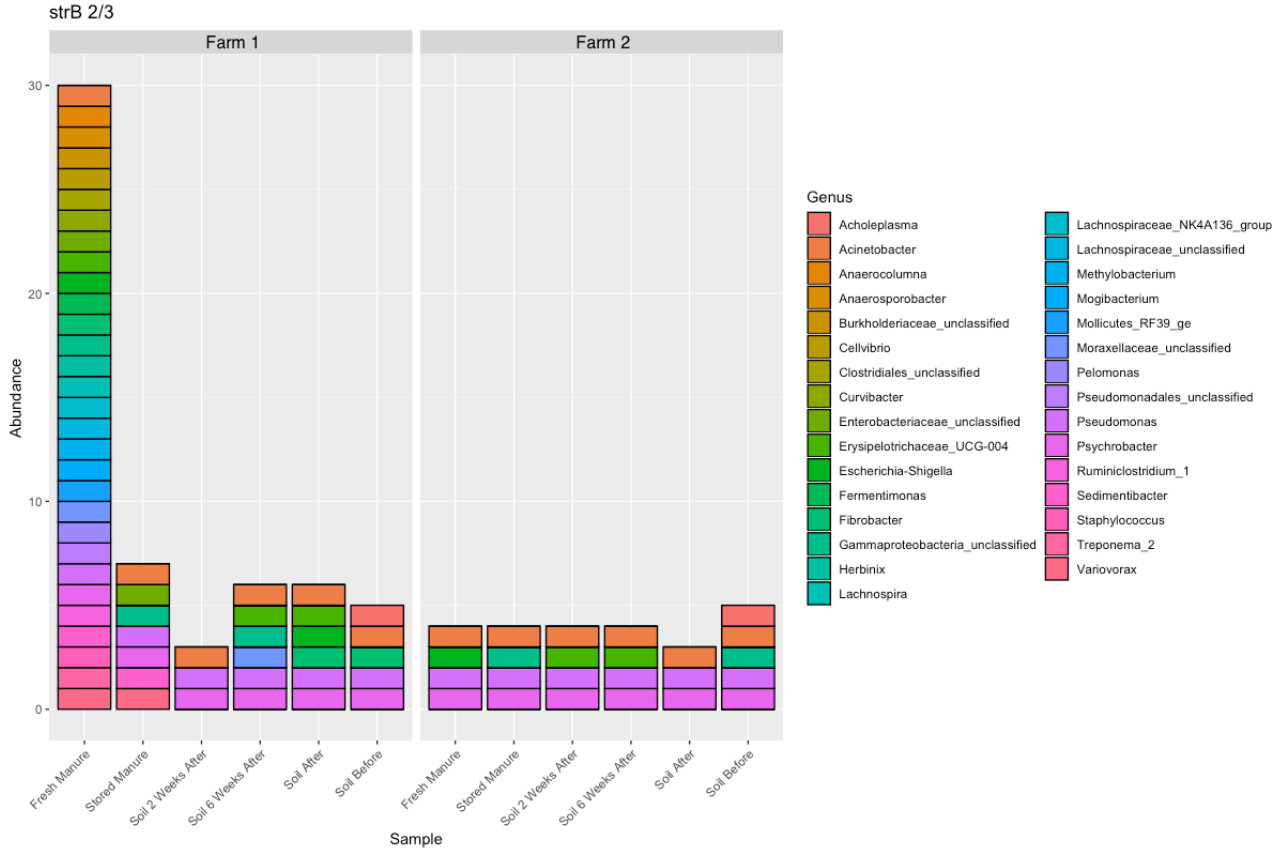

Figure S3. All detected OTUs present in *strB*

R script for 16S community analysis

```
library("phyloseq")
library("ggplot2")
library("vegan")
library("viridis")
library("ggrepel")
library("permute")
library("lattice")
library("tidyverse")
library("forcats")

setwd("~/Documents/16S_gradu")

#OTU table
OTU<-read.table("otutab_raw.txt", sep="\t", fill= 1, header=TRUE, row.names = 1)
#Taxonomy table
tax<-read.table("otus.taxonomy.txt", header=FALSE, sep="\t", row.names = 1)

#Sample data
sample_data <- read.table("sample_data.txt", sep="\t", header=TRUE, row.names= 1)

df <- as.data.frame(sample_data)
df$Sample_name <- rownames(sample_data)

#Merge into a phyloseq object
mydata<-phyloseq(otu_table((OTU), taxa_are_rows = TRUE), tax_table(as.matrix(tax)), sample_data(df))

###remove the confidence values assigned to each taxonomic level (the numbers inside parentheses)###
head(tax_table(mydata))
tax_table(mydata) <- gsub("\\([0-9]{1,3}\\)", "", tax_table(mydata))
head(tax_table(mydata))
colnames(tax_table(mydata))

#Change the column names to something meaningful in the taxonomy table
colnames(tax_table(mydata)) <- c("Domain", "Phylum", "Class", "Order", "Family", "Genus")
colnames(tax_table(mydata))
#Check sample sums
sample_sums(mydata)
mydata
head(tax_table(mydata))
sample_data(mydata)


# Check if there are any chloroplast sequences
subset_taxa(mydata, Class=="Chloroplast")
# And then remove them
mydata <- subset_taxa(mydata, Class!="Chloroplast")


#### Library sizes and normalization ####
lib_sizes <- sample_sums(mydata)
lib_sizes
barplot(lib_sizes)

mydata_prop <- transform_sample_counts(mydata, function(x) x/sum(x))
lib_sizes_prop <- sample_sums(mydata_prop)
lib_sizes_prop
barplot(lib_sizes_prop)

####ANNOTATION####
mydata_OTU <- as.data.frame(t(otu_table(mydata)))
adonis(mydata_OTU ~ ., sample_data, permutations = 99, method = "bray")

mydata_ord_pcoa <- dbrda(mydata_OTU ~ ., sample_data, distance = "bray")
```

```
plot(mydata_ord_pcoa)

mydata_prop_ord <- ordinate(mydata_prop, method="PCoA", distance="bray")

###principle cordinate analysis###
sample_data_ord <- ordinate(mydata, method = "PCoA", distance = "bray")

p <- plot_ordination(mydata, sample_data_ord, color = "SAMPLE.TYPE") + geom_point(size=3)

metaphlan.plot <- p + scale_color_manual(values=c("#E5FFCC", "black")) + geom_point(colour = "black",
                                                    pch = 21, size = 3, alpha = 0.5) + stat_ellipse(level = 0.95, linetype = 1) +
  theme_minimal() + labs(title = "16S")

metaphlan.plot

tmp_df <- data.frame("PCoA1"=data.frame(mydata_prop_ord$vectors)$Axis.1,
                "PCoA2"=data.frame(mydata_prop_ord$vectors)$Axis.2,
                SAMPLE.TYPE=sample_data(mydata_prop)$SAMPLE.TYPE, FARM=sample_data(mydata_prop)$FARM,
                SAMPLE_TYPE2 = sample_data(mydata_prop)$SAMPLE_TYPE2)

plot(sort(taxa_sums(mydata), decreasing = TRUE))

tmp_df

p<-ggplot(tmp_df, aes(x=PCoA1, y=PCoA2, color=SAMPLE.TYPE, shape=FARM, label=SAMPLE_TYPE2)) +
  geom_point(size = 3) +
  scale_color_manual(name="Sample", values=c("salmon4", "turquoise4")) +
  scale_shape(name="Farm") +
  stat_ellipse(aes(x=PCoA1, y=PCoA2, color=SAMPLE.TYPE), level = 0.95, linetype = 1, inherit.aes = FALSE) +
  theme_minimal() +
  xlab("PCoA 1, 63.5 %") +
  ylab("PCoA 2, 13.2 %") +
  theme(legend.position = "bottom")

p + geom_text_repel(color="black", segment.size  = 0.2,
                segment.color = "grey50",
                direction     = "y", hjust  = 2) + labs(title = "", color="black")


####Beta-diversity measured by jackard & bray###

adonis(t(otu_table(mydata_prop))~FARM+SAMPLE.TYPE, data=data.frame(sample_data(mydata)), permutations = 9999,
method="bray")
adonis(t(otu_table(mydata_prop))~FARM+SAMPLE.TYPE, data=data.frame(sample_data(mydata)), permutations = 9999,
binary=TRUE, method="jaccard")


mydata_prop_soil<-subset_samples(mydata_prop, SAMPLE_TYPE2%in%c("SA", "SB"))
adonis(t(otu_table(mydata_prop_soil))~SAMPLE_TYPE2, data=data.frame(sample_data(mydata_prop_soil)), permutations = 9999,
binary=TRUE, method="jaccard")
adonis(t(otu_table(mydata_prop_soil))~SAMPLE_TYPE2, data=data.frame(sample_data(mydata_prop_soil)), permutations = 9999,
method="bray")


mydata_prop_soil<-subset_samples(mydata_prop, SAMPLE_TYPE2%in%c("SB", "S6WA"))
adonis(t(otu_table(mydata_prop_soil))~SAMPLE_TYPE2, data=data.frame(sample_data(mydata_prop_soil)), permutations = 9999,
binary=TRUE, method="jaccard")
adonis(t(otu_table(mydata_prop_soil))~SAMPLE_TYPE2, data=data.frame(sample_data(mydata_prop_soil)), permutations = 9999,
method="bray")


mydata_prop_manure<-subset_samples(mydata_prop, SAMPLE_TYPE2%in%c("I", "M"))
adonis(t(otu_table(mydata_prop_manure))~SAMPLE_TYPE2, data=data.frame(sample_data(mydata_prop_manure)), permutations =
9999, binary=TRUE, method="jaccard")
adonis(t(otu_table(mydata_prop_manure))~SAMPLE_TYPE2, data=data.frame(sample_data(mydata_prop_manure)), permutations =
9999, method="bray")
```

```
mydata_prop_manure<-subset_samples(mydata_prop, SAMPLE.TYPE=="manure")
adonis(t(otu_table(mydata_prop_manure))~SAMPLE_TYPE2, data=data.frame(sample_data(mydata_prop_manure)), permutations =
9999, binary=TRUE, method="jaccard")
adonis(t(otu_table(mydata_prop_manure))~SAMPLE_TYPE2, data=data.frame(sample_data(mydata_prop_manure)), permutations =
9999, method="bray")


mydata_prop_soil<-subset_samples(mydata_prop, SAMPLE_TYPE2%in%c("SA", "S6WA"))
adonis(t(otu_table(mydata_prop_soil))~SAMPLE_TYPE2, data=data.frame(sample_data(mydata_prop_soil)), permutations = 9999,
binary=TRUE, method="jaccard")
adonis(t(otu_table(mydata_prop_soil))~SAMPLE_TYPE2, data=data.frame(sample_data(mydata_prop_soil)), permutations = 9999,
method="bray")


#### Plotting data ####

##Phylum level##
mydata_phylum <- tax_glom(mydata_prop, taxrank="Phylum")
plot_bar(mydata_phylum, fill="Phylum")

##Order level##
mydata_order <- tax_glom(mydata_prop, taxrank="Order")
mydata_order_abund <- prune_taxa(names(sort(taxa_sums(mydata_order),TRUE)[1:10]), mydata_order)

# Transfrom phyloseq object into dataframe
order_data = psmelt(mydata_order_abund)

# Barplot
ggplot(order_data, aes(x=Sample, y=Abundance)) +
  geom_bar(stat = "identity")

# Add color palette
order_figure <- ggplot(order_data, aes(x = Sample, y = Abundance, fill = Order)) +
  geom_bar(colour="black",stat="identity", width = NULL) +
  theme_minimal() +
  scale_fill_brewer(palette = "BrBG") +
  theme(legend.position="top")

order_figure

s <- order_data %>%
  mutate(Sample = fct_relevel(Sample,
                              "T1I", "T1M", "T1SB", "T1SA", "T1S2WA", "T1S6WA",
                              "T2I", "T2M", "T2SB", "T2SA", "T2S2WA", "T2S6WA")) %>%
  ggplot( aes(x=Sample, y=Abundance, fill =Order)) +
  geom_bar(colour="black",stat="identity") +
  xlab("Sample") +
  theme_minimal() +
  scale_fill_brewer(palette = "BrBG") +
  theme(legend.position="top")

s
s + facet_grid(~FARM, space="free", scales = "free",
               labeller = label_bquote(cols='Farm'~.(FARM)~"))  +ggtitle("Microbial Community") + scale_fill_brewer(palette="BrBG") +
  theme_minimal() + scale_x_discrete(labels = c('Fresh Manure','Stored Manure','Soil Before','Soil After','Soil 2 Weeks After','Soil 6
Weeks After',
                                      'Fresh Manure','Stored Manure','Soil Before','Soil After','Soil 2 Weeks After','Soil 6 Weeks After')) +
  theme(axis.text.x=element_text(angle=45, hjust=1))


#### Soil community plot #####
mydata_prop_soil<-subset_samples(mydata_prop, SAMPLE.TYPE=="soil")
mydata_soil_genus <- tax_glom(mydata_prop_soil, taxrank="Genus")
mydata_soil_genus_abund <- prune_taxa(names(sort(taxa_sums(mydata_soil_genus),TRUE)[1:10]), mydata_soil_genus)
```

```
(tax_table(mydata_soil_genus_abund))
tax_table(mydata_soil_genus_abund)[1,6]<-"Uncultured (Intrasporangiaceae)"
tax_table(mydata_soil_genus_abund)[3,6]<-"Candidatus Udaeobacter"
tax_table(mydata_soil_genus_abund)[5,6]<-"Uncultured (Xanthobacteraceae)"
tax_table(mydata_soil_genus_abund)[6,6]<-"Uncultured (Gaiellales)"
tax_table(mydata_soil_genus_abund)[7,6]<-"Candidatus Solibacter"
tax_table(mydata_soil_genus_abund)[8,6]<-"Uncultured (Gemmatimonadaceae)"
tax_table(mydata_soil_genus_abund)[9,6]<-"KD4-96_ge (Chloroflexi)"

# Transfrom phyloseq object into dataframe
genus_data = psmelt(mydata_soil_genus_abund)

# Barplot
ggplot(genus_data, aes(x=Sample, y=Abundance)) +
  geom_bar(stat = "identity")

# Add color palette
genus_figure <- ggplot(order_data, aes(x = Sample, y = Abundance, fill = "Genus")) +
  geom_bar(colour="black",stat="identity", width = NULL) +
  theme_minimal() +
  scale_fill_brewer(palette = "BrBG") +
  theme(legend.position="top")

genus_figure

g <- genus_data %>%
  mutate(Sample = fct_relevel(Sample,
                              "T1SB", "T1SA", "T1S2WA", "T1S6WA",
                              "T2SB", "T2SA", "T2S2WA", "T2S6WA")) %>%
  ggplot( aes(x=Sample, y=Abundance, fill =Genus)) +
  geom_bar(colour="black",stat="identity") +
  xlab("Sample") +
  theme_minimal() +
  scale_fill_brewer(palette = "BrBG") +
  theme(legend.position="top")

g
g + facet_grid(~FARM, space="free", scales = "free",
               labeller = label_bquote(cols='Farm'~.(FARM)~"))  +ggtitle("Soil Community") + scale_fill_brewer(palette="BrBG") +
  theme_minimal() + scale_x_discrete(labels = c('Soil Before','Soil After','Soil 2 Weeks After','Soil 6 Weeks After',
                                                'Soil Before','Soil After','Soil 2 Weeks After','Soil 6 Weeks After')) +
  theme(axis.text.x=element_text(angle=45, hjust=1))


##### Manure community plot #####
mydata_prop_manure<-subset_samples(mydata_prop, SAMPLE.TYPE=="manure")
mydata_manure_genus <- tax_glom(mydata_prop_manure, taxrank="Genus")
mydata_manure_genus_abund <- prune_taxa(names(sort(taxa_sums(mydata_manure_genus),TRUE)[1:10]), mydata_manure_genus)
(tax_table(mydata_manure_genus_abund))
tax_table(mydata_manure_genus_abund)[3,6]<-"Uncultured (Carnobacteriaceae)"
tax_table(mydata_manure_genus_abund)[5,6]<-"Ruminococcaceae"
tax_table(mydata_manure_genus_abund)[7,6]<-"Rikenellaceae"
plot_bar(mydata_manure_genus_abund, fill="Genus") + facet_grid(~FARM, space="free", scales = "free", labeller =
label_bquote(cols='Farm'~.(FARM)~")) +ggtitle("Manure Community") + scale_fill_brewer(palette="BrBG") + theme_minimal() +
scale_x_discrete(labels = c('Fresh Manure','Stored Manure','Fresh Manure','Stored Manure')) +
theme(axis.text.x=element_text(angle=45, hjust=1))

####See the percentage of how much different OTU's can be found from the samples####
manure_genera_percentage <- (data.frame(rowSums(100*(otu_table(mydata_manure_genus)))))
soil_genera_percentage <- (data.frame(rowSums(100*(otu_table(mydata_soil_genus)))))
View(soil_genera_percentage)
View(manure_genera_percentage)

# save percentage tables for soil and manure
write.table(soil_genera_percentage, "~/Documents/16S_gradu/soil_genera_percentage2.txt", sep="\t")
write.table(manure_genera_percentage, "~/Documents/16S_gradu/manure_genera_percentage2.txt", sep="\t")
```

R script for epicPCR analysis

```
library("phyloseq")
library("ggplot2")
library("decontam")
library("cowplot")
library("tidyverse")
library("forcats")

setwd("~/Documents/gradu2020")

#OTU table
OTU<-read.table("16S_epic_OTUs_test.txt", sep="\t", fill= 1, header=TRUE, row.names = 1)
#Taxonomy table
tax<-read.table("16S_epic.tax", header=FALSE, sep="\t", row.names = 1)
#Sample data
sample_data<-read.table("meta_data.txt", header=TRUE, sep="\t", row.names = 1)
#Merge into a phyloseq object
epic<-phyloseq(otu_table((OTU), taxa_are_rows = TRUE), tax_table(as.matrix(tax)), sample_data(sample_data))
#Change the column names to something meaningful in the taxonomy table
colnames(tax_table(epic)) <- c("Domain","Phylum", "Class", "Order", "Family", "Genus")
#Check sample sums
sample_sums(epic)

row.names(tax_table(epic))

#### tetM gene 2/3 #####

sample_data(epic)$is.neg<-sample_data(epic)$Sample_type%in%c("neg")
epic_tetM<-subset_samples(epic, Gene=="tetM")
contamdf.prev05 <- isContaminant(epic_tetM, method="prevalence", neg="is.neg", threshold=0.5)

epic_tetM_clean<-subset_taxa(epic_tetM, contamdf.prev05$contaminant==FALSE)

table(contamdf.prev05$contaminant)
head(which(contamdf.prev05$contaminant))

plot(sample_sums(epic_tetM_clean)~sample_data(epic_tetM_clean)$Sample_type)
sample_sums(epic_tetM_clean)

##### remove negative controls #####
epic_tetM_clean <- subset_samples(epic_tetM_clean, Sample_type!="neg")

plot(sample_sums(epic_tetM_clean)~sample_data(epic_tetM_clean)$Sample_type)
sample_sums(epic_tetM_clean)


#####Combine replicates######

sample_data(epic_tetM_clean)$Farm_type<-paste(sample_data(epic_tetM_clean)$Farm, sample_data(epic_tetM_clean)$Sample_type,
sep="_")
epic_tetM_mrg_clean<-merge_samples(epic_tetM_clean, "Farm_type", fun=sum)
sample_sums(epic_tetM_mrg_clean)
sample_data(epic_tetM_mrg_clean)


#Change to presence absence
epic_tetM_pres<-otu_table(epic_tetM_clean)
otu_table(epic_tetM_pres)[otu_table(epic_tetM_pres)>0]<-1
epic_tetM_pres<-phyloseq(otu_table(epic_tetM_pres), sample_data(sample_data), tax_table(as.matrix(tax_table(epic_tetM_clean))))
sample_data(epic_tetM_pres)$Farm_type<-paste(sample_data(epic_tetM_pres)$Farm, sample_data(epic_tetM_pres)$Sample_type,
sep="_")
#Merge biological replicates using Phyloseq
epic_tetM_mrg<-merge_samples(epic_tetM_pres, "Farm_type", fun=sum)

#Change all values less than 2 to 0 and more than 1 to  1 get presence absence data again
```

```
epic_tetM_mrg_pres<-otu_table(epic_tetM_mrg)
otu_table(epic_tetM_mrg_pres)[otu_table(epic_tetM_mrg_pres)<2]<-0
otu_table(epic_tetM_mrg_pres)[otu_table(epic_tetM_mrg_pres)>=2]<-1
epic_tetM_mrg_pres<-phyloseq(otu_table(epic_tetM_mrg_pres), sample_data(epic_tetM_mrg),
tax_table(as.matrix(tax_table(epic_tetM_clean))))

#### Genus level ####

epic_tetM_V2<-tax_glom(epic_tetM_mrg_pres, taxrank = "Genus")
otu_table(epic_tetM_V2)[otu_table(epic_tetM_V2)>0]<-1

# Take 12 most abundant
epic_V2_tetM_abun<-prune_taxa(names(sort(taxa_sums(epic_tetM_V2), TRUE)[1:12]), epic_tetM_V2)
(tax_table(epic_V2_tetM_abun))
tax_table(epic_V2_tetM_abun)[1,6]<-"VadinBB60 group (Clostridiales)"
tax_table(epic_V2_tetM_abun)[7,6]<-"NK4A136 group (Lachnospiraceae)"
tax_table(epic_V2_tetM_abun)[8,6]<-"Unclassified (Gammaproteobacteria)"
tax_table(epic_V2_tetM_abun)[2,6]<-"Unclassified (Enterobacteriaceae)"
tax_table(epic_V2_tetM_abun)[10,6]<-"Unclassified (Lachnospiraceae)"
tax_table(epic_V2_tetM_abun)[12,6]<-"Treponema"

# Transfrom phyloseq object into dataframe
tetM_data = psmelt(epic_V2_tetM_abun)

# Barplot
ggplot(tetM_data, aes(x=Sample, y=Abundance)) +
  geom_bar(stat = "identity")

# Add color palette
tetM_figure <- ggplot(tetM_data, aes(x = Sample, y = Abundance, fill = Genus)) +
  geom_bar(colour="black",stat="identity", width = NULL) +
  theme_minimal() +
  scale_fill_brewer(palette = "Set3") +
  theme(legend.position="top")

tetM_figure

t <- tetM_data %>%
  mutate(Sample = fct_relevel(Sample,
                              "T1_I", "T1_M", "T1_SB", "T1_SA", "T1_S2WA", "T1_S6WA",
                              "T2_I", "T2_M", "T2_SB", "T2_SA", "T2_S2WA", "T2_S6WA")) %>%
  ggplot( aes(x=Sample, y=Abundance, fill =Genus)) +
  geom_bar(colour="black",stat="identity") +
  xlab("") +
  theme_minimal() +
  scale_fill_brewer(palette = "Set3") +
  theme(legend.position="top")

t
t + facet_grid(~Farm, space="free", scales = "free", labeller = label_bquote(cols="")) +ggtitle("tetM 2/3")


#### combining reads to ARG barplot ####

sample_sums(epic_tetM_mrg_clean)
df <- data.frame(sum=sample_sums(epic_tetM_mrg_clean),
                 sample_type=sample_names(epic_tetM_mrg_clean), Farm=c(rep("Farm 1", 6), rep("Farm 2", 6)))

df

head(df)
a <- df %>%
  mutate(sample_type = fct_relevel(sample_type,
                              "T1_I", "T1_M", "T1_SB", "T1_SA", "T1_S2WA", "T1_S6WA",
                              "T2_I", "T2_M", "T2_SB", "T2_SA", "T2_S2WA", "T2_S6WA")) %>%
  ggplot( aes(x=sample_type, y=sum)) + facet_grid(~Farm, space="free", scales = "free") +
```

```
  geom_bar(stat="identity", fill="aquamarine4") + theme_minimal() + geom_area() + scale_x_discrete(labels = c('Fresh Manure','Stored Manure','Soil Before','Soil After','Soil 2 Weeks After','Soil 6 Weeks After',
                                                                                                              'Fresh Manure','Stored Manure','Soil Before','Soil After','Soil 2 Weeks After','Soil 6 Weeks After')) + theme(axis.text.x=element_text(angle=45, hjust=1))

a + scale_y_sqrt(breaks=c(1000, 10000, 20000, 30000, 40000, 50000, 60000, 70000, 80000)) + scale_y_reverse()



plot_grid(t +labs(title="Host range of tetM", x="", y = "Number of hosts") + theme_minimal() + theme(panel.spacing = unit(1, "lines"))
        + theme(legend.position = "top") + theme(axis.text.x = element_blank(), axis.ticks = element_blank())
        + scale_fill_brewer(palette="Set3") + scale_y_continuous(breaks=c(2,4,6,8,10)), a
        + scale_y_sqrt(breaks=c(1000, 10000, 20000, 30000, 40000, 50000, 60000, 70000, 80000))
        +labs(title="Library size", x="Sample", y = "Sum") + theme(legend.position = "top")
        + theme(axis.ticks = element_blank()) + scale_y_reverse()
        + geom_text(aes(label=sum), vjust=1.2, color="black", size=3.5), nrow = 2, labels="AUTO", align = "v")




#### blaOXA58 2/3 ####

sample_data(epic)$is.neg<-sample_data(epic)$Sample_type%in%c("neg")
epic_blaOXA58<-subset_samples(epic, Gene=="blaOXA58")
contamdf.prev05 <- isContaminant(epic_blaOXA58, method="prevalence", neg="is.neg", threshold=0.1)

epic_blaOXA58_clean<-subset_taxa(epic_blaOXA58, contamdf.prev05$contaminant==FALSE)

plot(sample_sums(epic_blaOXA58_clean)~sample_data(epic_blaOXA58_clean)$Sample_type)
sample_sums(epic_blaOXA58_clean)

##### remove
epic_blaOXA58_clean <- subset_samples(epic_blaOXA58_clean, Sample_type!="neg")

plot(sample_sums(epic_blaOXA58_clean)~sample_data(epic_blaOXA58_clean)$Sample_type)
sample_sums(epic_blaOXA58_clean)

#### Combine replicates #####

sample_data(epic_blaOXA58_clean)$Farm_type<-paste(sample_data(epic_blaOXA58_clean)$Farm,
sample_data(epic_blaOXA58_clean)$Sample_type, sep="_")
epic_blaOXA58_mrg_clean<-merge_samples(epic_blaOXA58_clean, "Farm_type", fun=sum)
sample_sums(epic_blaOXA58_mrg_clean)


#Change to presence absence
epic_blaOXA58_pres<-otu_table(epic_blaOXA58_clean)
otu_table(epic_blaOXA58_pres)[otu_table(epic_blaOXA58_pres)>0]<-1
epic_blaOXA58_pres<-phyloseq(otu_table(epic_blaOXA58_pres), sample_data(sample_data),
tax_table(as.matrix(tax_table(epic_blaOXA58_clean))))
sample_data(epic_blaOXA58_pres)$Farm_type<-paste(sample_data(epic_blaOXA58_pres)$Farm,
sample_data(epic_blaOXA58_pres)$Sample_type, sep="_")
#Merge biological replicates using Phyloseq
epic_blaOXA58_mrg<-merge_samples(epic_blaOXA58_pres, "Farm_type", fun=sum)
sample_data(epic_blaOXA58_mrg)$Farm[(sample_data(epic_blaOXA58_mrg)$Farm==2)]<-3


#Change all values less than 2 to 0 and more than 1 to  1 get presence absence data again
epic_blaOXA58_mrg_pres<-otu_table(epic_blaOXA58_mrg)
otu_table(epic_blaOXA58_mrg_pres)[otu_table(epic_blaOXA58_mrg_pres)<2]<-0
otu_table(epic_blaOXA58_mrg_pres)[otu_table(epic_blaOXA58_mrg_pres)>=2]<-1
epic_blaOXA58_mrg_pres<-phyloseq(otu_table(epic_blaOXA58_mrg_pres), sample_data(epic_blaOXA58_mrg),
tax_table(as.matrix(tax_table(epic_blaOXA58_clean))))

#### Genus level ####
```

```
epic_blaOXA58_V2<-tax_glom(epic_blaOXA58_mrg_pres, taxrank = "Genus")
otu_table(epic_blaOXA58_V2)[otu_table(epic_blaOXA58_V2)>0]<-1


#Take 12 most abundant
epic_V2_blaOXA58_abun<-prune_taxa(names(sort(taxa_sums(epic_blaOXA58_V2), TRUE)[1:7]), epic_blaOXA58_V2)
(tax_table(epic_V2_blaOXA58_abun))
tax_table(epic_V2_blaOXA58_abun)[1,6]<-"Unclassified (Enterobacteriaceae)"
tax_table(epic_V2_blaOXA58_abun)[5,6]<-"Unclassified (Lachnospiraceae)"

# Transfrom phyloseq object into dataframe
blaOXA58_data = psmelt(epic_V2_blaOXA58_abun)

# Barplot
ggplot(blaOXA58_data, aes(x=Sample, y=Abundance)) +
  geom_bar(stat = "identity")

# Add color palette
blaOXA58_figure <- ggplot(blaOXA58_data, aes(x = Sample, y = Abundance, fill = Genus)) +
  geom_bar(colour="black",stat="identity", width = NULL) +
  theme_minimal() +
  scale_fill_brewer(palette = "Set3") +
  theme(legend.position="top")

blaOXA58_figure

b <- blaOXA58_data %>%
  mutate(Sample = fct_relevel(Sample,
                              "T1_I", "T1_M", "T1_SB", "T1_SA", "T1_S2WA", "T1_S6WA",
                              "T2_I", "T2_M", "T2_SB", "T2_SA", "T2_S2WA", "T2_S6WA")) %>%
  ggplot( aes(x=Sample, y=Abundance, fill =Genus)) +
  geom_bar(colour="black",stat="identity") +
  xlab("") +
  theme_minimal() +
  scale_fill_brewer(palette = "Set3") +
  theme(legend.position="top")

b
b + facet_grid(~Farm, space="free", scales = "free", labeller = label_bquote(cols="")) +ggtitle("blaOXA58 2/3")


#### combining reads to ARG barplot ####

sample_sums(epic_blaOXA58_mrg_clean)
df <- data.frame(sum=sample_sums(epic_blaOXA58_mrg_clean),
                 sample_type=sample_names(epic_blaOXA58_mrg_clean), Farm=c(rep("Farm 1", 6), rep("Farm 2", 6)))

df

head(df)
c <- df %>%
  mutate(sample_type = fct_relevel(sample_type,
                                   "T1_I", "T1_M", "T1_SB", "T1_SA", "T1_S2WA", "T1_S6WA",
                                   "T2_I", "T2_M", "T2_SB", "T2_SA", "T2_S2WA", "T2_S6WA")) %>%
  ggplot( aes(x=sample_type, y=sum)) + facet_grid(~Farm, space="free", scales = "free") +
  geom_bar(stat="identity", fill="aquamarine4") + theme_minimal() + geom_area() + scale_x_discrete(labels = c('Fresh Manure','Stored
Manure','Soil Before','Soil After','Soil 2 Weeks After','Soil 6 Weeks After',
                                                                                'Fresh Manure','Stored Manure','Soil Before','Soil After','Soil 2
Weeks After','Soil 6 Weeks After')) + theme(axis.text.x=element_text(angle=45, hjust=1))

c + scale_y_sqrt(breaks=c(100, 1000, 2000, 3000, 4000, 5000, 6000)) + scale_y_reverse()
```

```
plot_grid(b +labs(title="Host range of blaOXA58", x="", y = "Number of hosts") + theme_minimal() + theme(panel.spacing = unit(1,
"lines"))
        + theme(legend.position = "top") + theme(axis.text.x = element_blank(), axis.ticks = element_blank())
        + scale_fill_brewer(palette="Set3") + scale_y_continuous(breaks=c(2,4,6,8,10)), c
        + scale_y_sqrt(breaks=c(100, 1000, 2000, 3000, 4000, 5000, 6000))
        +labs(title="Library size", x="Sample", y = "Sum") + theme(legend.position = "top")
        + theme(axis.ticks = element_blank()) + scale_y_reverse()
        + geom_text(aes(label=sum), vjust=1.2, color="black", size=3.5), nrow = 2, labels="AUTO", align = "v")


#### strB gene 2/3 ####

sample_data(epic)$is.neg<-sample_data(epic)$Sample_type%in%c("neg")
epic_strB<-subset_samples(epic, Gene=="strB")
contamdf.prev05 <- isContaminant(epic_strB, method="prevalence", neg="is.neg", threshold=0.1)

epic_strB_clean<-subset_taxa(epic_strB, contamdf.prev05$contaminant==FALSE)

plot(sample_sums(epic_strB_clean)~sample_data(epic_strB_clean)$Sample_type)
sample_sums(epic_strB_clean)

##### remove
epic_strB_clean <- subset_samples(epic_strB_clean, Sample_type!="neg")

plot(sample_sums(epic_strB_clean)~sample_data(epic_strB_clean)$Sample_type)
sample_sums(epic_strB_clean)

######Combine replicates #####

sample_data(epic_strB_clean)$Farm_type<-paste(sample_data(epic_strB_clean)$Farm, sample_data(epic_strB_clean)$Sample_type,
sep="_")
epic_strB_mrg_clean<-merge_samples(epic_strB_clean, "Farm_type", fun=sum)
sample_sums(epic_strB_mrg_clean)


#Change to presence absence
epic_strB_pres<-otu_table(epic_strB_clean)
otu_table(epic_strB_pres)[otu_table(epic_strB_pres)>0]<-1
epic_strB_pres<-phyloseq(otu_table(epic_strB_pres), sample_data(sample_data), tax_table(as.matrix(tax_table(epic_strB_clean))))
sample_data(epic_strB_pres)$Farm_type<-paste(sample_data(epic_strB_pres)$Farm, sample_data(epic_strB_pres)$Sample_type,
sep="_")

#Merge biological replicates using Phyloseq
epic_strB_mrg<-merge_samples(epic_strB_pres, "Farm_type", fun=sum)

#Change all values less than 2 to 0 and more than 1 to  1 get presence absence data again
epic_strB_mrg_pres<-otu_table(epic_strB_mrg)
otu_table(epic_strB_mrg_pres)[otu_table(epic_strB_mrg_pres)<2]<-0
otu_table(epic_strB_mrg_pres)[otu_table(epic_strB_mrg_pres)>=2]<-1
epic_strB_mrg_pres<-phyloseq(otu_table(epic_strB_mrg_pres), sample_data(epic_strB_mrg),
tax_table(as.matrix(tax_table(epic_strB_clean))))


epic_strB_V2<-tax_glom(epic_strB_mrg_pres, taxrank = "Genus")
otu_table(epic_strB_V2)[otu_table(epic_strB_V2)>0]<-1


#Take 12 most abundant
epic_V2_strB_abun<-prune_taxa(names(sort(taxa_sums(epic_strB_V2), TRUE)[1:12]), epic_strB_V2)
tax_table(epic_V2_strB_abun)[11,6]<-"Unclassified (Moraxellaceae)"
tax_table(epic_V2_strB_abun)[1,6]<-"Unclassified (Enterobacteriaceae)"
tax_table(epic_V2_strB_abun)[7,6]<-"Unclassified (Gammaproteobacteria)"
tax_table(epic_V2_strB_abun)[8,6]<-"Unclassified (Erysipelotrichaceae)"

# Transfrom phyloseq object into dataframe
strB_data = psmelt(epic_V2_strB_abun)
```

```
# Barplot
ggplot(strB_data, aes(x=Sample, y=Abundance)) +
  geom_bar(stat = "identity")

# Add color palette
strB_figure <- ggplot(strB_data, aes(x = Sample, y = Abundance, fill = Genus)) +
  geom_bar(colour="black",stat="identity", width = NULL) +
  theme_minimal() +
  scale_fill_brewer(palette = "Set3") +
  theme(legend.position="top")

strB_figure

s <- strB_data %>%
  mutate(Sample = fct_relevel(Sample,
                              "T1_I", "T1_M", "T1_SB", "T1_SA", "T1_S2WA", "T1_S6WA",
                              "T2_I", "T2_M", "T2_SB", "T2_SA", "T2_S2WA", "T2_S6WA")) %>%
  ggplot( aes(x=Sample, y=Abundance, fill =Genus)) +
  geom_bar(colour="black",stat="identity") +
  xlab("") +
  theme_minimal() +
  scale_fill_brewer(palette = "Set3") +
  theme(legend.position="top")

s
s + facet_grid(~Farm, space="free", scales = "free", labeller = label_bquote(cols="")) +ggtitle("strB 2/3")


#### combining reads to ARG barplot ####

sample_sums(epic_strB_mrg_clean)
df <- data.frame(sum=sample_sums(epic_strB_mrg_clean),
                 sample_type=sample_names(epic_strB_mrg_clean), Farm=c(rep("Farm 1", 6), rep("Farm 2", 6)))

df

head(df)
g <- df %>%
  mutate(sample_type = fct_relevel(sample_type,
                                   "T1_I", "T1_M", "T1_SB", "T1_SA", "T1_S2WA", "T1_S6WA",
                                   "T2_I", "T2_M", "T2_SB", "T2_SA", "T2_S2WA", "T2_S6WA")) %>%
  ggplot( aes(x=sample_type, y=sum)) + facet_grid(~Farm, space="free", scales = "free") +
  geom_bar(stat="identity", fill="aquamarine4") + theme_minimal() + geom_area() + scale_x_discrete(labels = c('Fresh Manure','Stored
Manure','Soil Before','Soil After','Soil 2 Weeks After','Soil 6 Weeks After',
                                                                                'Fresh Manure','Stored Manure','Soil Before','Soil After','Soil 2
Weeks After','Soil 6 Weeks After')) + theme(axis.text.x=element_text(angle=45, hjust=1))

g + scale_y_sqrt(breaks=c(1000, 10000, 20000, 30000, 40000, 50000, 60000, 70000, 80000, 90000, 100000, 200000, 300000,
350000)) + scale_y_reverse()




plot_grid(s +labs(title="Host range of strB", x="", y = "Number of hosts") + theme_minimal() + theme(panel.spacing = unit(1, "lines"))
        + theme(legend.position = "top") + theme(axis.text.x = element_blank(), axis.ticks = element_blank())
        + scale_fill_brewer(palette="Set3") + scale_y_continuous(breaks=c(2,4,6,8,10)), g
        + scale_y_sqrt(breaks=c(1000, 10000, 20000, 30000, 40000, 50000, 60000, 70000, 80000, 90000, 100000, 200000, 300000,
350000))
        +labs(title="Library size", x="Sample", y = "Sum") + theme(legend.position = "top")
        + theme(axis.ticks = element_blank()) + scale_y_reverse()
        + geom_text(aes(label=sum), vjust=1.2, color="black", size=3.5), nrow = 2, labels="AUTO", align = "v")
```